

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2004, 19:03:15 ; Search time 21 Seconds

(without alignments)
892,995 Million cell updates/sec

Title: US-09-997-701-1

Perfect score: 1110
Sequence: 1 MESWMLPCLAFPLCFMHAR.....SYFKLNRRNCPTTHPEPV 195Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR 76:*

1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1004.5	90.5	180	2 S43791	PBX protein - hum
2	160.5	14.5	185	2 A60592	T-cell surface gly
3	146	13.2	415	2 F89994	hypothetical prote
4	139	12.5	527	2 B70700	hypothetical prote
5	136.5	12.3	435	2 T15143	hypothetical prote
6	135.5	12.2	300	2 JQ2220	hydroxyproline-ric
7	128	11.5	348	1 OZQ2BK	circumsporozoite p
8	127.5	11.5	707	2 A46302	PTB-associated spl
9	125	11.3	332	1 OZQ2BK	circumsporozoite p
10	125	11.3	1585	2 T13611	hypothetical prote
11	124	11.2	3164	1 NMBE66	Urb6 protein - hum
12	123.5	11.1	177	2 S65780	glycine/proline-ri
13	123.5	11.1	300	2 S19560	proline-rich prote
14	123.5	11.1	378	2 S14959	proline-rich prote
15	122	11.0	301	2 E29149	proline-rich prote
16	121.5	10.9	407	2 T12956	hypothetical prote
17	121.5	10.9	463	2 T10015	hypothetical prote
18	121.5	10.9	488	2 F86911	conserved hypotet
19	121.5	10.9	633	2 F84564	probable protein k
20	121	10.9	676	2 S41022	hypothetical prote
21	120	10.8	350	2 I55214	salivary proline-r
22	118.5	10.7	478	2 T21144	hypothetical prote
23	117	10.5	350	2 T36611	probable integral
24	117	10.5	922	2 S15204	pectactin - bordet
25	116.5	10.5	437	2 T26767	hypothetical prote
26	116.5	10.5	500	2 T20961	hypothetical prote
27	116.5	10.5	530	2 T48627	hypothetical prote
28	116.5	10.5	940	2 J80291	FB19 protein - hum
29	116	10.5	245	1 M4WLB5	E4 protein - human

30	116	10.5	900	2 B70694	probable infR - My
31	115.5	10.4	161	2 S12246	another-specific pr
32	115	10.4	264	2 A54330	major prion protei
33	115	10.4	289	2 F96770	protein RNA-bindin
34	115	10.4	545	2 T15667	hypothetical prote
35	115	10.4	924	2 F87103	initiation factor
36	114.5	10.3	228	2 S46955	microfilament shea
37	114.5	10.3	387	2 T21370	hypothetical prote
38	114.5	10.3	505	2 T50815	cdc2-like protein
39	114.5	10.3	641	2 JC6017	zinc-finger transc
40	114.5	10.3	798	2 T21369	hypothetical prote
41	114.5	10.3	2715	2 T13049	eyelid - fruit fly
42	114	10.3	571	2 T43456	hypothetical prote
43	114	10.3	730	2 T43317	p91-1 protein - Ca
44	114	10.3	771	2 T29177	hypothetical prote
45	113.5	10.2	260	2 S22373	proline-rich prote

ALIGNMENTS

RESULT 1

S43791
PBX protein - human

C:Species: Homo sapiens (man)

C:Date: 13-Jan-1995 #sequence_revision 10-Nov-1995 #text_change 07-May-1999

C:Accession: S43791

R:Ellis, N.A.; Ye, T.Z.; Patton, S.; German, J.; Goodfellow, P.N.; Weller, P.

Nature Genet. 6, 394-399, 1994

A:Title: Cloning of PBX, an M2-related gene that spans the pseudautosomal boundary

A:Reference number: S43791; MUID:94332149; PMID:8054981

A:Accession: S43791

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-180 <ELL>

A:Cross-references: GB:S73261; NID:9639633; PID:9639634

Query Match 90.5%; Score 1004.5; DB 2; Length 180;
Best Local Similarity 92.3%; Pred. No. 6.56-66;
Matches 180; Conservative 0; Mismatches 0; Indels 15; Gaps 1;

QY	1	MESWMLPCLAFPLCFMHARGQRPDLADLDDDEPTKKPSDLYPKPKPYQPEPPD	60
DB	1	MESWMLPCLAFPLCFMHARGQRPDLADLDDDEPTKKPSDLYPKPKPYQPEPPD	60
QY	61	SGGNIYRPRKPRPQPOGNSGSGYFNDVDRDGRYPAPRPAPAGGGGGYSSYGS	120
DB	61	SGGNIYRPRKPRPQPOGNSGSGYFNDVDRDGRYPAPRPAPAGGGGGYSSYGS	120
QY	121	DNTHRGQGRPNRSRYGNTYGGDHHSTYGNPEGNVAKIVSPISVVVTLIGAAASYFKL	180
DB	121	DNTHRGQGRPNRSRYGNTYGGDHHSTYGNPEGNVAKIVSPISVVVTLIGAAASYFKL	180
QY	181	NNRNCPTTHPEPV 195	
DB	166	NNRNCPTTHPEPV 180	

RESULT 2

A60592
T-cell surface glycoprotein E2 precursor - human

N:Alternate names: 32k protein; MIT2 antigen

C:Species: Homo sapiens (man)

C:Date: 17-Apr-1993 #sequence_revision 03-Feb-1994 #text_change 05-Nov-1999

C:Accession: S06786; A32876; A60592

R:Gelein, C.; Abrit, F.; Phalipon, A.; Raynal, B.; Cole, S.; Kaczorek, M.; Bernard, A.

EMBO J. 8, 3253-3259, 1989

A:Title: The E2 antigen, a 32 kd glycoprotein involved in T-cell adhesion processes, is

A:Reference number: S06786; MUID:9005916; PMID:2479542

A:Accession: S06786

A:Molecule type: mRNA

A:Residues: 1-185 <GEI>

A:Cross-references: EMBL:X16996; NID:930948; PIND:CAA34863.1; PID:930949

Query Match	14.5%	Score 160.5	DB 2	Length 185
Best Local Similarity	28.0%	Pred. No. 6, 2e-05		
Matches	56	Conservative 16	Mismatches 57	Indels 71
			Gaps 8	
QY	10	LALFLC---LMAHAGQRFDLADLDDPEPTKKGNSDIYPKKPPY-----	53	
Db	7	LALLFLGLGLVLAAPDGGFDLSDALPUNE-NKKPTA-----IPKKPSAGDDFDLDGDAVD	61	
QY	54	-----POPENPDSGANIYPRPKRRPOQPGNSGNSGGYFNVVDYRDGGYRPPRRPRRPRAG	108	
Db	62	GENDDPRPPNP-----PKMENPNP-NHPSSGSTSDADLDG-----VSG	101	
QY	109	GGGGGSSSYGNSDNTHGRGGYRPNRYGNTYGGDHHSTYGNPBGNNVAIVSPIVSVVVV	168	
Db	102	GEGKGGSGGGSHRKEGEBAAPG-----VIGIVGAVVV	136	
QY	169	TLLGAAAYFKLNNRNCFR	188	
Db	137	AVAGAISSFIAVQKQCLCFK	156	

```

Query Match      13.2% Score 146; DB 2; Length 415;
Best Local Similarity 27.5%; Pwd. No. 0.0015;
Matches 46; Conservative 19; Mismatches 58; Indels 44; Gaps 7;

09      33 DPEPKKNSDLYPKPKP-PYRPOB-----IPDSGNTYRPKRPPQPPQNSGNSGGY 86
      :||| ||: | ||||| | |: ||: |:||| | :| | :

```

RESULT 4
B70700
hypochemical protein RV0020C - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: B70700
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
J.; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentile, S.; Hamlin, N.; Holtrold, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellon, S.; Squares, S.
Nature 393. 537-544, 1998
A:Authors: Squares, R.; Sulstrom, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:96295987; PMID:9634230
A:Accession: B70700
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-527 <COL>
A:Cross-references: GB:Z80233; GB:AL123456; NID:g3261645; PIDD:CAB02440.1; PID:e2669377;
A:Experimental source: strain H37RV
C:Genetics:
I:Gene: RV0020C

RESULT 5
T15143
hypothetical protein T28F2.8 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jan-2000
C:Accession: T15143
R:MadSEN, C.; Fromick, B.
submitted to the EMBL Data Library, April 1997
A:Description: The sequence of *C. elegans* cosmid T28F2.
A:Reference number: Z18300
A:Accession: T15143
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-435 <MMD>
A:Cross-references: EMBL:AF000198, NID:g2047345, PID:g2047353, PIND:AAB51059.1, GSPDB:C:Caenorhabditis elegans Bristol N2; clone T28F2
A:Experimental source: strain Bristol N2; clone T28F2

A;Gene: CESP:T28F2.8
A;Map position: 1
C;Superfamily: Phaeococcus glycine-rich cell wall protein 1.8
Query March 12.3%; Score 136.5; DB 2; Length 435;
Best Local Similarity 37.9%; Pred. No. 0.0078;

Matches 50; Conservative 3; Mismatches 54; Indels 25; Gaps 8;

QY DPEP-----TKFENSIITYKPK-PPYYPQPEPND-SGNITYPKPRPOP-----QPGNSG 81

Db DQGPAPAPGNTNVSNGCGGPGAGPPGPPGPGDGGGAQCGPPGPPGPPGPNDCQPGGPG 293

QY 82 NSGGFEND-VBRDDGRYPFRPRPPAGGGGGGYSYNSNDNTGRCGYRPNRSRYGNTY 140

Db 294 QPGGGGQGGGERTDAAYCPCP-PRTPAGGGGGGDFPAG-----GGGSY-----STGG 340

QY 141 GDHHSYGNPBG 152

Db 341 GGGRADSGGAAG 352

RESULT 6

QJ02220

hydroxyproline-rich glycoprotein precursor - Chlamydomonas reinhardtii

C.Species: Chlamydomonas reinhardtii

C.Date: 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change 21-Jul-2000

C.Accession: U02220

R.Malfen Schmid, S.; Woessner, J.P.; Beer, K.; Goodenough, U.W.

Plant Cell 5, 809-820, 1993

A.Title: Isodityrosine cross-linking mediates insolubilization of cell walls in chlamydomonas

A.Reference number: U02220; MUID:93372571; PMID:7689882

A.Accession: U02220

A.Molecule type: mRNA

A.Residues: 1-300 <MAR>

A.Cross-references: GB:L16461; NID:g269960; PID:AAA02923.1; PID:g269961

A.Experimental source: cDNA: VSP-1

C.Comment: This protein is a component of the insoluble inner layer W2 of vegetative/gam

C.Keywords: cell wall, glycoprotein

F.1-28/Domain: signal sequence #status predicted <SIG>

F.129-300/Product: hydroxyproline-rich glycoprotein #status predicted <MAT>

F.187-189,197-205,211-212,222-241/Region: 3-residue repeats (Y-G-G)

Query Match 12.2%; Score 135.5; DB 2; Length 300;

Best local similarity 30.7%; Pred. No. 0.0064;

Matches 39; Conservative 7; Mismatches 40; Indels 41; Gaps 7;

QY 32 DDPEPTKPNSDIYPKPPYYPQPEPNDSGNITYPKPRPOP-----GNSGN 82

Db 142 DSPSPSPPEVEDSPSPPE--PMDESP-----APSPSPPEVSPETPPSPATYGAAPS 193

QY 83 SGGFENDVDRDDGRYP-----RPRPPAGGGGGGYSYNSNDNTGRCGYRPNRSRY 135

Db 194 PSPYGYG-----GNVGFPEVPTYGSSPSPAPPGYGGYGYG-----GYGGY----- 237

QY 136 GNTYGGD 142

Db 238 -GSYGDD 243

RESULT 7

OZ020K

circumsporozoite protein precursor - Plasmodium berghei (strain ANKA clone 2.34L)

N.Alternate names: sporozoite surface antigen

C.Species: Plasmodium berghei

C.Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jul-1999

C.Accession: S07873; S12571

R.Lockyer, M.J.; Davies, C.S.; Subbier, A.; Sinden, R.E.

Nucleic Acids Res. 18, 376, 1990

A.Title: Nucleotide sequence of the Plasmodium berghei circumsporozoite protein gene

A.Reference number: S07873; MUID:90221834; PMID:2183186

A.Accession: S07873

A.Molecule type: DNA

A.Residues: 1-348 <LOC>

A.Cross-references: EMBL:X17606

R.Lockyer, M.J.

submitted to the EMBL Data Library, November 1989

A.Reference number: S12571

A.Accession: S12571

A.Molecule type: DNA

[illegible]

Query Match 11.1%; Score 123.5; DB 2; Length 177;
 Best Local Similarity 31.9%; Pred. No. 0.028;
 Matches 45; Conservative 4; Mismatches 43; Indels 49; Gaps 9;

QY 45 YPK---PKPYYPQENPDGSGNITYPKPKPPQPPQPGNSGSGYFNDVDDDRYPRP 101
 Db 16 YPPAGYPPPGAYPPAGYPOG---YP-PPGAYPPAG-----YPPGAYPPAP 58
 QY 102 RPPPPAGGGG---GYSYSGNSDNTGRGGRPNRYGNTYGGDHRSTYGNPQNVVA 156
 Db 59 GGYPPAPGYGYPAPAGYGGYPPAP---GHGGYPPAG---YPA-HHSHGAGGIGMIA 109
 QY 157 KIVSPYIVSVVVTLLGAASY 177
 Db 110 -----GAAAY 115

RESULT 13

S19560
 C/Species: Mus musculus (house mouse)
 C/Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 20-Aug-1999
 C/Accession: S19560; S22570
 R:Roberts, S.G.E.; Layfield, R.; Bannister, A.J.; McDonald, C.J.
 A/Title: Gene sequence of mouse B-type proline-rich protein MP4. Transcriptional start P
 A/Reference number: S19560; MUID:92111548; PMID:1765104
 A/Accession: S19560
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-300 <ROB>
 A/Cross-references: GB:X58438; NID:G53181; PIDN:CAA41344.1; PID:G53182
 R:Roberts, S.G.E.; Layfield, R.; McDonald, C.J.
 A/Title: The mouse proline-rich protein MP6 promoter binds isoprenaline-inducible parotid
 A/Reference number: S22570; MUID:92020206; PMID:1747160
 A/Accession: S22570
 A/Molecule type: DNA
 A/Residues: 1-14 <RO2>
 A/Cross-references: EMBL:X61126
 C/Superfamily: proline-rich protein

Query Match 11.1%; Score 123.5; DB 2; Length 300;
 Best Local Similarity 31.2%; Pred. No. 0.047;
 Matches 53; Conservative 6; Mismatches 58; Indels 53; Gaps 11;

QY 18 HARQGRDFDLADLDDPEPTKPKPNSDIYPKPKPPY-----POPEN---PDGSGNIX 66
 Db 93 HAPPOQ---GDALGPPR---GNQGGPSPGPPQSSSQRRPQPGNQGPPOGG--- 142
 QY 67 PRPKRPQPPQGN-----SGNSGYFNDVDRDDGRYPRPRP-----R 104
 Db 143 --PQRRP-PQPGNQGPSPGPPQRRPQPGNQG-----PQGGPHRP-PRRGNQGP 193
 QY 105 PPAGGGGGGYSYSGNSDNTGRGGRPNRYGNTYGGDHRSTYGNP-GN 153
 Db 194 PPQGGPQRRPQPGNQGPPOGGPQAPRRPQPGNQGPSPGPPQGPRTGN 243

RESULT 14

S14959
 C/Species: Triticum aestivum (common wheat)
 C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 24-Nov-1999
 C/Accession: S14959
 R:Raines, C.A.; Lloyd, J.C.; Chao, S.; John, U.P.; Murphy, G.J.P.
 A/Title: A novel proline-rich protein from wheat.
 A/Reference number: S14959; MUID:91329699; PMID:1714320
 A/Accession: S14959
 A/Status: preliminary
 A/Molecule type: mRNA

A/Residues: 1-378 <RAI>
 A/Cross-references: EMBL:X52472; NID:G21841; PIDN:CAA36712.1; PID:G21842
 C/Superfamily: hydroxyproline-rich glycoprotein

Query Match 11.1%; Score 123.5; DB 2; Length 378;
 Best Local Similarity 28.6%; Pred. No. 0.059;
 Matches 24; Conservative 15; Mismatches 28; Indels 17; Gaps 2;

QY 29 DALDDEPTKPKPNSDIYPKPKPPYPOPE---NPDGSGNITYPKPKPPQPPQPGNSGSG 85
 Db 302 EPMKPEPKPEPKPEPMKPEPKPEPKPEPKPEPKPEPKPEPKPEPKPEPKPEPKPE 353
 QY 86 YFNDVDRDDGRYPRPRPPAG 109
 Db 354 -----KPEPKPEPKPEPKPEPKG 371

RESULT 15

E29149
 C/Species: Mus musculus (house mouse)
 C/Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 20-Aug-1999
 C/Accession: E29149
 R:Clements, S.; Mehansho, H.; Carlson, D.M.
 A/Title: Novel multigene families encoding highly repetitive peptide sequences. Sequence
 A/Reference number: A92501; MUID:86033799; PMID:3840480
 A/Contents: Clone pUMPI25
 A/Accession: E29149
 A/Molecule type: mRNA
 A/Residues: 1-301 <CLB>
 A/Cross-references: GB:M11997; NID:G200540; PIDN:AAA40001.1; PID:G200541
 C/Superfamily: proline-rich protein

Query Match 11.0%; Score 122; DB 2; Length 301;
 Best Local Similarity 28.8%; Pred. No. 0.06;
 Matches 45; Conservative 9; Mismatches 52; Indels 50; Gaps 7;

QY 34 PEPTKPKPNSDIYPKPKPPYQENPDGSGNITYPKPKPPQ---PQPGNS----- 80
 Db 148 PQGNQGG-----PPPGPPQRRPQPGNQGPSPGPPRGNQGPSPGPPQ 201
 QY 81 -----GNSGYFNDVDRDDGRYPR-----PRP-----PPAGGGGGYS 115
 Db 202 RPTQPGNQG---PQGGPQGPSPRGNQGCCPPGPGQGPSPRGNQGPSPGPG 258
 QY 116 SYGNSDNTGRGGRPNRYGNTYGGDHRSTYGNP 151
 Db 259 RPNQGGPSPGSPQRRPQPGNHQGPGR---GNNE 291

Search completed: February 18, 2004, 19:07:04
 Job time: 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: February 18, 2004, 18:56:46 ; Search time 17 Seconds

(without alignments)
539,424 Million cell updates/sec

Title: US-09-997-701-1

Perfect score: 1110
Sequence: 1 MESMGJLPCIAFLCFLMHAR.....SYFKLNRRNCFTREBENV 195

Scoring table: BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1004.5	90.5	180	1	XG_HUMAN
2	160.5	14.5	185	1	MIC2_HUMAN
3	128	11.5	347	1	CSP_PLAAR
4	127.5	11.5	707	1	SFPQ_HUMAN
5	125	11.3	339	1	CSP_PLAAR
6	124.5	11.2	3164	1	TEGU_HSV1
7	119	10.7	245	1	VEA_HPVS
8	118	10.6	1790	1	SEPA_EMENI
9	117	10.5	922	1	PERT_BORPA
10	116	10.5	245	1	VEA_HPVS
11	116	10.5	900	1	IF2_MYCTU
12	115.5	10.4	161	1	ASF1_HUMAN
13	115.5	10.4	264	1	PRIO_BOVIN
14	115	10.4	924	1	IF2_MYCTU
15	113.5	10.2	1664	1	SEPI_CLOTH
16	113	10.2	824	1	TEM1_RAT
17	112.5	10.1	255	1	PRIO_CANFA
18	112.5	10.1	704	1	VEA_HPVS
19	112	10.1	304	1	VEA_HPVS
20	112	10.1	1059	1	CAPU_DROME
21	112	10.1	448	1	AMC2_DICDI
22	111.5	10.0	143	1	PAR1_TRYBB
23	110.5	10.0	145	1	PARC_TRYBB
24	110.5	10.0	257	1	PRIO_MUSPF
25	110.5	10.0	257	1	PRIO_MUSPF
26	110.5	10.0	283	1	EXTN_SORBI
27	110.5	10.0	283	1	EXTN_SORBI
28	110.5	10.0	345	1	SOD_DROME
29	110	9.9	911	1	IF3_MOUSE
30	109.5	9.9	911	1	COTI_BACSU
31	109.5	9.9	449	1	APG_BRANA
32	109	9.8	449	1	PRP2_BOVIN
33	109	9.8	305	1	ROAD_HUMAN

34	109	9.8	373	1	RO31_XENLA
35	108.5	9.8	129	1	PARB_TRYBB
36	108.5	9.8	174	1	PRP2_HUMAN
37	108.5	9.8	256	1	PRIO_CEREL
38	108.5	9.8	526	1	FUS_HUMAN
39	108.5	9.8	534	1	APG_ARATH
40	108.5	9.8	625	1	DUS8_HUMAN
41	108	9.7	256	1	PRIO_CAPI
42	108	9.7	256	1	PRIO_FELCA
43	108	9.7	256	1	PRIO_SHEEP
44	108	9.7	514	1	MEP_MOUSE
45	107.5	9.7	512	1	FUS_BOVIN

ALIGNMENTS

RESULT 1
XG_HUMAN STANDARD; PRT; 180 AA.

AC P55808;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE XG glycoprotein precursor (Protein PBDX).
GN XG OR PBDX.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Bone marrow;
RC MEDLINE=94332149; Pubmed=8054981;
RA Ellis N.A., Ye T.Z., Patton S., German J., Goodfellow P.N.,
RA Waller P.,
RT "Cloning of PBDX, an MIC2-related gene that spans the pseudautosomal
RT boundary on chromosome Xp.";
RL Nat. Genet. 6:394-400(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Moko S., Fujiki K., Kanai A., Tanaka Y., Iwata T.,
RT "Identification of PBDX gene highly expressed in human cornea.";
RL Submitted (May-2001) to the EMBL/Genbank/DBJ database.
CC - SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC - PTM: O-Glycosylated (PROBABLY).
CC - POLYMORPHISM: XG IS RESPONSIBLE FOR THE XG BLOOD GROUP SYSTEM.
CC - SIMILARITY: TO PROTEIN MIC2/CD99.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: X96421; -; NOT ANNOTATED CDS.
DR EMBL: AF80356; AAL04055.1; -;
DR PIR: S43791; S43791.
DR GeneW: HGNC:12806; XG.
DR MIM: 314700; -;
KW Transmembrane; Glycoprotein; Blood group antigen; signal.
FT SIGNAL 1 21
FT CHAIN 22 180 XG GLYCOPROTEIN.
FT DOMAIN 22 142 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 143 163 POTENTIAL.
FT DOMAIN 164 180 CYTOPLASMIC (POTENTIAL).
SQ SSOURCE 180 AA; 19723 MW; DADA95859C4530 CRC64;
Query Match 90.5%; Score 1004.5; DB 1; Length 180;
Best Local Similarity 92.3%; Pred. No. 4.3e-62;
Matches 180; Conservative 0; Mismatches 0; Indels 15; Gaps 1;

```

QY 1 MESWGLPLCLAFCLFMAHARGRDFDLADLDDPEPTKXKNSDIYKPKPPYPPEND 60
DB 1 MESWGLPLCLAFCLFMAHARGRDFDLADLDDPEPTKXKNSDIYKPKPPYPPEND 60
QY 61 SGNITPRKPRPOPOPGNSGGYFNDVDRDGRYPRPRPPRPPAGGGGCGSYSSYNS 120
DB 61 SGNITPRKPRPOPOPGNSGGYFNDVDRDGRYPRPRPPRPPAGGGGCGSYSSYNS 120
QY 121 DNTGRGGYRPNRYGNTYGGDHSHTYGNPEGNMVAKIVSPYVVVVTLLGAASYEKL 180
DB 121 DNTH-----GGDHSHTYGNPEGNMVAKIVSPYVVVVTLLGAASYEKL 165
QY 181 NNRNRCFRTHEPENV 195
DB 166 NNRNRCFRTHEPENV 180

RESULT 2
MIC2_HUMAN STANDARD; PRT; 185 AA.
AC P14209; 000518;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE T-cell surface glycoprotein E2 precursor (E2 antigen) (CD99 antigen)
DE (MIC2 protein) (12E7).
GN (CD99X OR CD99 OR MIC2X OR MIC2) AND (CD99Y OR CD99 OR MIC2Y OR MIC2).
OC Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE OF 23-39.
RC TISSUE=T-cell;
RX MEDLINE=90059916; PubMed=2479542;
RA Gellin C., Aubric F., Phalipon A., Raynal B., Cole S., Kaczorek M.,
RA Bernard A.;
RT "The E2 antigen, a 32 kd glycoprotein involved in T-cell adhesion
RT processes, is the MIC2 gene product."
RT EMBO J. 8:3253-3259(1989).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM II).
RC Park S.H., Hahn J.H., Kim M.K., Sohn H.W., Choi E.Y., Kim S.H.;
RX "An alternative splicing form of CD99 (MIC2).";
RT Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 10-105 FROM N.A.
RX MEDLINE=87217010; PubMed=3472717;
RA Darling S.M., Goodfellow P.J., Pym B., Bantling G.S., Pritchard C.,
RA Goodfellow P.N.;
RT "Molecular genetics of MIC2: a gene shared by the human X and Y
RT chromosomes."
RL Cold Spring Harb. Symp. Quant. Biol. 51:205-212(1986).
RN [4]
RP SEQUENCE OF 1-22 FROM N.A.
RX MEDLINE=88289765; PubMed=2456574;
RA Goodfellow P.J., Mondello C., Darling S.M., Pym B., Little P.,
RA Goodfellow P.N.;
RT "Absence of methylation of a CpG-rich region at the 5' end of the
RT MIC2 gene on the active X, the inactive X, and the Y chromosome."
RL Proc. Natl. Acad. Sci. U.S.A. 85:5605-5609(1988).
RN [5]
RP FUNCTION: INVOLVED IN T-CELL ADHESION PROCESSES. IT IS INVOLVED IN
RN SPONTANEOUS ROSETTE FORMATION WITH ERYTHROCYTES.
CC [6]
RP SUBCELLULAR LOCATION: Type I membrane protein (potential).
CC [7]
RP ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=P14209-1; Sequence=Displayed;
CC Name=11;
CC IsoId=P14209-2; Sequence=VSP_004324;
CC [8]
RP PTM: EXTENSIVELY O-GLYCOSYLATED.
CC [9]
RP SIMILARITY: TO PROTEIN XG/PBDX.

```

```

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; X16996; CA34863.1; -
DR EMBL; U82164; AAB58501.1; -
DR EMBL; M16279; AAA02999.1; -
DR EMBL; J03841; AAA59848.1; -
DR PIR; S06786; A60592.
DR Genew; HGNC:7082; CD99.
DR MIM; 450000; -
DR GO; GO:0005737; C:cytoplasm; TAS.
DR GO; GO:0005887; C:integral to plasma membrane; TAS.
DR GO; GO:0005888; C:protein integral to plasma membrane; TAS.
KW T-cell; Glycoprotein; Cell adhesion; Transmembrane; Signal;
KW Alternative splicing; Polymorphism.
FT SIGNAL 1 22
FT CHAIN 23 185 T-CELL SURFACE GLYCOPROTEIN E2.
FT TRANSMEM 123 147 POTENTIAL.
FT VAAPSLIC 159 185 ABOGVMESEHNNANAEPAVORTILEK -> DG (in
FT isoform II).
FT VARIANT 166 166 M -> V (IN dbSNP:4793).
FT VARIANT 173 173 N -> I (IN dbSNP:4717).
FT FTId=VAR_014734.
FT FTId=VAR_014734.
SQ SEQUENCE 185 AA; 18848 MW; C302E09EBB022EAB CRC64;

Query Match 14.5%; Score 160.5; DB 1; Length 185;
Best Local Similarity 28.0%; Pred. No. 0.0001;
Matches 56; Conservative 16; Mismatches 57; Indels 71; Gaps 8;

QY 10 LAFICF---LMARGRDFDLADLDDPEPTKXKNSDIYKPKPPY----- 53
DB 7 LALLFGLGLVAPAPGGFDLADLDDPEPTKXKNSDIYKPKPPY----- 61
QY 54 -----POPNPDSGNITPRKPRPOPOPGNSGGYFNDVDRDGRYPRPRPPAG 108
DB 62 GENDPPEPNP-----PKPWPNNP-NHPSGSGFSFDADLADG-----VSG 101
QY 109 GGGGYSYSGNSDNTGRGGYRPNRYGNTYGGDHSHTYGNPEGNMVAKIVSPYVVVV 168
DB 102 GEGKGGSDGGGSHKKEBEADAPG-----VIRGIVGVAVV 136
QY 169 TLGAASYEKLNNRNCFR 188
DB 137 AVAGAISSFIAYOKKCLCFK 156

RESULT 3
CSP_PLABA STANDARD; PRT; 347 AA.
AC P23093;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Circumsporozoite protein precursor (CS).
OS Plasmodium berghei (strain Anka).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxId=5823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90221834; PubMed=2183186;
RA Lockyer M.J., Davies C.S., Sunbrier A., Sinden R.E.;
RT "Nucleotide sequence of the Plasmodium berghei circumsporozoite
RT protein gene from the ANKA clone 2.34L."
RL Nucleic Acids Res. 18:376-376(1990).

```

CC -1- FUNCTION: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT
 CC SURFACE ANTIGEN ON THE SPROZOITE (THE INFECTIVE STAGE OF THE
 CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE
 CC VERTEBRATE HOST).
 CC -1- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR
 CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES
 CC WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.
 CC -1- SIMILARITY: Contains 1 TSP type-1 domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X17606; CAA35608.1; -.
 CC PIR: S07873; OZQOBK.
 CC InterPro: IPR003067; Crcmaprozoite.
 CC Pfam: PF00090; tsp_1; 1.
 CC PRINTS: PR01303; CRCMSPROZOITE.
 CC SMART: SM00209; TSP1.1.
 CC PROSITE: PS50092; TSP1.1.
 CC Malaria; Sporozoite; Repeat; Signal.
 CC SIGNAL 1 23 POTENTIAL.
 CC CHAIN 24 347 CIRCUMSPOROZOITE PROTEIN.
 CC DOMAIN 93 204 13 X 8 AA REPEATS.
 CC DOMAIN 214 247 17 X 2 AA REPEATS OF P-Q.
 CC DOMAIN 274 325 TSP TYPE-1.
 CC SEQUENCE 347 AA; 37776 MW; DEC240EE35681AF8 CRC64;
 CC
 CC Query Match 11.5%; Score 128; DB 1; Length 347;
 CC Best Local Similarity 30.8%; Pred. No. 0.03;
 CC Matches 33; Conservative 15; Mismatches 33; Indels 24; Gaps 4;
 CC
 CC 32 DDEPTKKNSDIYKRP-----PYYPQENP-DSGNTYPRKPRPQPGNS 80
 CC 171 NDPPP-----PNANDPPPNPNDPAPNANDPPPNPNDPAPQGNNDPQPPQPP--- 224
 CC
 CC 81 GNSGGRYNDVDRDGRPRPRPRPRPAGGCGGCGSSGNSDNTH 124
 CC 225 -----QPQPQPQPQPQPQPQPQPQPQGNNNNNKNNNDST 261
 CC
 CC RESULT 4
 CC SFPQ_HUMAN STANDARD; PRT; 707 AA.
 CC ID SFPQ_HUMAN STANDARD; PRT; 707 AA.
 CC AC P33246; F30808;
 CC DT 01-NOV-1991 (Rel. 20, Created)
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE Splicing factor, proline- and glutamine-rich [Poly(pyrimidine tract-
 CC binding protein-associated splicing factor) (PTB)-associated tract-
 CC binding factor] (PSF) (DNA-binding P52/P100 complex, 100 kDa subunit).
 CC DE SFPQ OR PSF.
 CC OS Homo sapiens (Human).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CC OX NCBI_TaxId=9606;
 CC RN [1]
 CC RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ALTERNATIVE SPLICING.
 CC RC TISSUE=Fetal brain;
 CC RX MEDLINE=93194059; PubMed=8449401;
 CC RA Paton J.G., Porro E.B., Galceran J., Tempst P., Nadal-Ginard B.;
 CC RT "Cloning and characterization of PSF, a novel pre-mRNA splicing
 CC factor.";
 CC RL Genes Dev. 7:393-406(1993).
 CC RN [2]
 CC RP SEQUENCE OF 312-707 FROM N.A.
 CC RC TISSUE=Fetal skeletal muscle;
 CC RX MEDLINE=90091812; PubMed=2480877;

RA Gower H.J., Moore S.E., Dickson G., Elsom V.L., Nayak R., Walsh F.S.;
 CC "Cloning and characterization of a myoblast cell surface antigen
 CC RT defined by 24.1D5 monoclonal antibody.";
 CC RL Development 105:723-731(1989).
 CC
 CC RN [3]
 CC RP SEQUENCE OF 48-68 AND 213-246.
 CC RX MEDLINE=93176127; PubMed=8439294;
 CC RA Zhang W.-W., Zhang L.-X., Busch R.K., Farres J., Busch H.;
 CC RT "Purification and characterization of a DNA-binding heterodimer of 52
 CC RT and 100 kDa from HeLa cells.";
 CC RL Biochem. J. 290:267-272(1993).
 CC
 CC -1- FUNCTION: ESSENTIAL PRE-mRNA SPLICING FACTOR REQUIRED EARLY IN
 CC SPLICEOSOME FORMATION. BINDS TO THE MAMMALIAN POLYPYRIMIDINE
 CC TRACTS. FORMS A COMPLEX WITH THE POLYPYRIMIDINE TRACT-BINDING
 CC PROTEIN (PTB). SEEMS TO ALSO BIND DNA.
 CC -1- SUBUNIT: HETEROTRIMER OF TWO 52 KDA AND TWO 100 KDA SUBUNITS.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=Long;
 CC IsoId=P23246-1; Sequence=Displayed;
 CC Name=Short;
 CC IsoId=P23246-2; Sequence=VSP_005855;
 CC
 CC -1- SIMILARITY: Contains 2 RNA recognition motif (RRM) domains.
 CC -1- CAUTION: WAS ORIGINALLY (REF.2) THOUGHT TO BE MYOBLAST CELL
 CC SURFACE ANTIGEN 24.1D5 AND A POSSIBLE MEMBRANE-BOUND PROTEIN
 CC ECTOKINASE.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X70944; CAA50283.1; -.
 CC EMBL: X16850; CAA34747.1; -.
 CC PIR: A46302; A46302.
 CC HSSP: P11940; 1CVJ.
 CC SWISS-2DPAGE; P23246; HUMAN.
 CC Genew; HGNC:10774; SFPQ.
 CC DR MIM; 605199; -.
 CC DR GO; GO:0008248; P:pre-mRNA splicing factor activity; TAS.
 CC DR GO; GO:0006371; P:mRNA splicing; TAS.
 CC DR InterPro; IPR000504; RNA_rec_mot.
 CC DR Pfam; PF00076; Rrm_2.
 CC DR SMART; SM00360; RRM; 2.
 CC DR PROSITE; PS50102; RRM; 2.
 CC DR PROSITE; PS00030; RRM_RNP_1; 1.
 CC KW Nuclear protein; RNA-binding; DNA-binding; mRNA splicing; Repeat;
 CC Alternative splicing
 CC FT DOMAIN 297 369 RNA-BINDING (RRM) 1.
 CC FT DOMAIN 371 452 RNA-BINDING (RRM) 2.
 CC FT REPEAT 9 27 1. X 3 AA REPEATS OF R-G-G.
 CC FT REPEAT 19 21 1.
 CC FT REPEAT 25 27 2.
 CC FT DOMAIN 10 266 GLN/GLU/PRO-RICH.
 CC FT DOMAIN 10 15 POLY-GLY.
 CC FT DOMAIN 20 27 POLY-GLY.
 CC FT DOMAIN 56 65 POLY-PRO.
 CC FT DOMAIN 67 71 POLY-GLN.
 CC FT DOMAIN 95 98 POLY-GLN.
 CC FT DOMAIN 99 103 POLY-PRO.
 CC FT DOMAIN 184 188 POLY-PRO.
 CC FT DOMAIN 571 574 POLY-ARG.
 CC FT DOMAIN 613 616 POLY-GLY.
 CC FT DOMAIN 635 641 POLY-GLY.
 CC FT VARSPLIC 663 707
 CC
 CC RTERGGGAGVGGGPRGPGMGPTAGYGRGREGYEGPNK
 CC KPRF -> VRMIDVG (in isoform Short).

FT CONFLICT 243 /FTID=VSP_005855.
SQ SEQUENCE 707 AA; 76149 MW; 6DBD5E95E235847 CRC64;
Matches 42; Conservative 5; Mismatches 57; Indels 43; Gaps 6;

Query Match
Best Local Similarity 11.5%; Score 127.5; DB 1; Length 707;
Matches 42; Conservative 5; Mismatches 57; Indels 43; Gaps 6;

34 PEPTCKPNSDI-----YKPKPPYYPQENDPSGNN-----IYRPRPRQPOQNSG 81
147 PGPTLPPTAVTSAPGAPAPPTPSSGVTTTPQAGPPPPPAVPGPGPKGPGPGG 206
QY 82 NSGGYFNDVDRDGGYPPRRPPRPPAGGGGGSSYSYSGNDTHGGGTRPNRSRYNTYGG 141
DB 207 PKG-----GKMPGGPKP-----GGGPGISTPGHPKPPHRRGGSPRG-----GR 245
QY 142 DHHSY-----GNPEGMMVAKI 158
DB 246 QHPPYHQHOGPPPGGPGGSEKI 272

RESULT 5
CSP_PLABE STANDARD; PRT; 339 AA.
AC P06S15;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Circumsporozoite protein precursor (CS).
OS Plasmodium berghei.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NX NCBI_TaxID=5821;
RN [1]
RP MEDLINE=87089740; PubMed=2432395;
RA Eichinger D.J., Annot D.E., Tam J.P., Nussenzweig V., Enea V.;
RT "Circumsporozoite protein of Plasmodium berghei: gene cloning and
RT identification of the immunodominant epitopes."
RL Mol. Cell. Biol. 6:3965-3972(1986).
CC -1- FUNCTION: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT
CC SURFACE ANTIGEN ON THE SPOROZOITE (THE INFECTIVE STAGE OF THE
CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE
CC VERTEBRATE HOST).
CC -1- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR
CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES
CC WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.
CC -1- SIMILARITY: Contains 1 TSP type-1 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; M14135; AAA29577.1; -
DR EMBL; A44948; OZQMB.
DR InterPro; IPR003067; CrcmSprzoite.
DR Pfam; PIR000864; TSP1.
DR Pfam; PR00090; tsp.1; 1.
DR PRINTS; PR01303; CRCMSPRZOITE.
DR SMART; SM00209; TSP1; 1.
DR PROSITE; PS50092; TSP1; 1.
DR Malaria; Sporozoite; Repeat; 1.
FT SIGNAL 1 23 PROBABLE.
FT CHAIN 24 339 CIRCUMSPOROZOITE PROTEIN.
FT DOMAIN 93 196 13 X 8 AA TANDEM REPEATS.
FT DOMAIN 206 238 16 X 2 AA TANDEM REPEATS OF P-Q.
FT DOMAIN 266 317 TSP TYPE-1.
SQ SEQUENCE 339 AA; 37138 MW; E8068A6D1D9551B CRC64;
Query Match 11.3%; Score 125; DB 1; Length 339;

Best Local Similarity 32.3%; Pred. No. 0.048;
Matches 32; Conservative 15; Mismatches 30; Indels 22; Gaps 5;

QY 32 DDEPTCKPNSDIYKPKPPY--PQENDPS-----GNTYRPRKPPQPOQNSGSGG 85
DB 171 NDPAF---PNAN---DAPPNANDPPPNNDPAPPGGNNNPQPPPPQPP----- 216
QY 86 YFNDVDRDGGYPPRRPPRPPAGGGGGSSYSYSGNDTH 124
DB 217 --QP 253

RESULT 6
ID DRPL_RAT STANDARD; PRT; 1183 AA.
AC P54258;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Atrophin-1 (dentatorubral-pallidolysian atrophy protein).
GN DRPLA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum, and Striatum;
RX MEDLINE=97317138; PubMed=9173996;
RA Loev S.J., Margolis R.L., Young W.S., Li S.-H., Schilling G.,
RA Ashworth R.G., Ross C.A.;
RT "Cloning and expression of the rat atrophin-1 (DRPLA disease gene)
RT homologue."
RL Neurobiol. Dis. 2:129-138(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, Cerebellum, Hippocampus, and Substantia nigra;
RX MEDLINE=96081227; PubMed=8541849;
RA Schmitt I., Epplen J.T., Riese O.;
RT "Predominant neuronal expression of the gene responsible for
RT dentatorubral-pallidolysian atrophy (DRPLA) in rat."
RL Hum. Mol. Genet. 4:1619-1624(1995).
CC -1- TISSUE SPECIFICITY: PREDOMINANT NEURONAL EXPRESSION, ALTHOUGH
CC MARKEDLY REDUCED AMOUNTS ARE FOUND IN MOST OTHER TISSUES.
CC -1- DEVELOPMENTAL STAGE: SIMILAR EXPRESSION AT ALL DEVELOPMENT STAGES
CC (DAY 14.5 P.C., 17.5 P.C., NEWBORNS AND ADULTS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; U31777; AAA80337.1; -
DR EMBL; X89453; CA61623.1; -
DR InterPro; IPR002951; Atrophin.
DR Pfam; PF03154; Atrophin-1; 2.
DR PRINTS; PR01222; ATROPHIN.
FT DOMAIN 165 171 POLY-PRO.
FT DOMAIN 303 306 POLY-PRO.
FT DOMAIN 377 383 POLY-SER.
FT DOMAIN 387 391 POLY-SER.
FT DOMAIN 440 446 POLY-PRO.
FT DOMAIN 477 480 POLY-HIS.
FT DOMAIN 481 489 POLY-GLN.
FT DOMAIN 502 505 POLY-PRO.
FT DOMAIN 562 572 POLY-SER.
FT DOMAIN 702 705 POLY-PRO.
FT DOMAIN 707 705 POLY-SER.
FT CONFLICT 455 455
FT CONFLICT 594 594
FT CONFLICT 689 689
P -> R (IN REF. 2).
P -> R (IN REF. 2).

CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 CC Eurotiaceae; Trichocomaceae; Emmentella.
 CC NCBI_TaxID=162425;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC MEDLINE=97361829; PubMed=9218790;
 CC Harris S.D., Hamer L., Sharpless K.E., Hamer J.E.;
 CC "The Aspergillus nidulans sepa gene encodes an Fhl/2 protein involved
 CC in cytokinesis and the maintenance of cellular polarity.";
 CC EMO 16:3474-3483(1997).
 CC [2]
 CC REVISIONS TO 143-153; 207; 1071-1109 AND 1644.
 CC Hamer L., Harris S.D., Sharpless K.E., Hamer J.E.;
 CC Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
 CC [3]
 CC SEQUENCE OF 559-1790 FROM N.A., AND FUNCTION.
 CC STRAIN=FGSC 26;
 CC MEDLINE=95229045; PubMed=7713416;
 CC Martoul J.F., Adams T.H.;
 CC "Identification of developmental regulatory genes in Aspergillus
 CC nidulans by overexpression.";
 CC Genetics 139:537-547(1995).
 CC -1- FUNCTION: INVOLVED IN CYTOKINESIS. OVEREXPRESSION RESULTS IN
 CC GROWTH INHIBITION.
 CC -1- DOMAIN: DFRS ARE REGULATED BY INTRAMOLECULAR GBD-DAD BINDING WHERE
 CC RHO-GTP ACTIVATES THE DFRS BY DISRUPTING THE GBD-DAD INTERACTION
 CC (BY SIMILARITY).
 CC -1- SIMILARITY: Contains 1 GTPase-binding (GBD) domain.
 CC -1- SIMILARITY: Contains 1 Formin homology 2 (FH2) domain.
 CC -1- SIMILARITY: Contains 1 Formin homology 3 (FH3) domain.
 CC -1- SIMILARITY: Contains 1 DRF autoregulatory (DAD) domain.
 CC -1- SIMILARITY: BELONGS TO THE FORMIN HOMOLOG FAMILY. EN11
 CC SUPFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC DR EMBL; U83558; AAB63335.3; -; ALT_SEQ.
 CC DR EMBL; L36341; AAA33306.1; -; ALT_SEQ.
 CC DR InterPro; IPR003104; FH2; 1.
 CC DR Pfam; PF02181; FH2; 1.
 CC DR SMART; SM00498; FH2; 1.
 CC Cell division; Coiled coil.
 CC KW DOMAIN 258 486 GBD.
 CC FT DOMAIN 376 718 FH3.
 CC FT DOMAIN 724 811 COILED COIL (POTENTIAL).
 CC FT DOMAIN 955 1136 FH1 (PRO-RICH).
 CC FT DOMAIN 1141 1658 FH2.
 CC FT DOMAIN 1435 1566 COILED COIL (POTENTIAL).
 CC FT DOMAIN 1586 1600 DAD.
 CC FT DOMAIN 1608 1611 AGG/lys-RICH (BASIC).
 CC FT DOMAIN 1071 1109 PPPPPGFGFGPPPPGFGPPPPGFGPPPPGFG
 CC CONFLICT 1071 1109 -> LSPSSGTGFWMTTPAASPTWFMSTAAASYMWC
 CC C (IN REF. 3).
 CC FT CONFLICT 1476 1476 D -> V (IN REF. 3).
 CC FT CONFLICT 1504 1504 V -> L (IN REF. 3).
 CC FT SEQUENCE 1790 AA; 197355 MW; 192136DE2A75B CRC64;
 CC -----
 CC Query Match 10.6%; Score 118; DB 1; Length 1790;
 CC Best Local Similarity 35.2%; Pred. No. 0.69; Indels 22; Gaps 3;
 CC Matches 32; Conservative 0; Mismatches 37;
 CC -----
 CC QY 26 DLADLADDEPTKCN---SDIYKPKPYTPQENPDGSGNITPRKRPQPPGNSG 81
 CC DB 1010 DAKTAAAPPPPPPPHPLSGAAPP---PPPPPPGAGAAPPPPPPPPGGLG 1066
 CC QY 82 NSGGYFNDVDRDGRPRPRPPAGGGG 112

CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 CC Eurotiaceae; Trichocomaceae; Emmentella.
 CC NCBI_TaxID=162425;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC MEDLINE=97361829; PubMed=9218790;
 CC Harris S.D., Hamer L., Sharpless K.E., Hamer J.E.;
 CC "The Aspergillus nidulans sepa gene encodes an Fhl/2 protein involved
 CC in cytokinesis and the maintenance of cellular polarity.";
 CC EMO 16:3474-3483(1997).
 CC [2]
 CC REVISIONS TO 143-153; 207; 1071-1109 AND 1644.
 CC Hamer L., Harris S.D., Sharpless K.E., Hamer J.E.;
 CC Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
 CC [3]
 CC SEQUENCE OF 559-1790 FROM N.A., AND FUNCTION.
 CC STRAIN=FGSC 26;
 CC MEDLINE=95229045; PubMed=7713416;
 CC Martoul J.F., Adams T.H.;
 CC "Identification of developmental regulatory genes in Aspergillus
 CC nidulans by overexpression.";
 CC Genetics 139:537-547(1995).
 CC -1- FUNCTION: INVOLVED IN CYTOKINESIS. OVEREXPRESSION RESULTS IN
 CC GROWTH INHIBITION.
 CC -1- DOMAIN: DFRS ARE REGULATED BY INTRAMOLECULAR GBD-DAD BINDING WHERE
 CC RHO-GTP ACTIVATES THE DFRS BY DISRUPTING THE GBD-DAD INTERACTION
 CC (BY SIMILARITY).
 CC -1- SIMILARITY: Contains 1 GTPase-binding (GBD) domain.
 CC -1- SIMILARITY: Contains 1 Formin homology 2 (FH2) domain.
 CC -1- SIMILARITY: Contains 1 Formin homology 3 (FH3) domain.
 CC -1- SIMILARITY: Contains 1 DRF autoregulatory (DAD) domain.
 CC -1- SIMILARITY: BELONGS TO THE FORMIN HOMOLOG FAMILY. EN11
 CC SUPFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC DR EMBL; X54547; CA38419.1; -;
 CC DR EMBL; A26124; CA01786.1; -;
 CC DR EMBL; A19182; CA01454.1; -;
 CC DR PIR; S15204; S15204.
 CC DR InterPro; IPR006315; Autotransport.
 CC DR InterPro; IPR005546; Autotransporter.
 CC DR InterPro; IPR004899; Pertactin.
 CC DR InterPro; IPR003991; Pertactin_C.
 CC DR InterPro; IPR003992; Pertactin_N.
 CC Pfam; PF03797; Autotransporter; 1.
 CC Pfam; PF03212; Pertactin; 1.
 CC PRINTS; PRO1482; PERTACTIN.
 CC DR PRINTS; PRO1484; PERTACTINFAMILY.
 CC DR TIGRFAMs; TIGR01414; autocrans Bar1; 1.
 CC KW Outer membrane; Signal; Virulence; Repeat.
 CC FT SIGNAL 1 34 POTENTIAL.
 CC FT CHAIN 35 922 P. 95.
 CC FT CHAIN 35 647 POTENTIAL.
 CC FT PROPEP 648 922 POTENTIAL.
 CC FT SITE 260 922 CELL ATTACHMENT SITE (INVOLVED IN
 CC ADHESION TO VARIOUS EUKARYOTIC CELL
 CC LINES).
 CC FT DOMAIN 266 290 4 X 5 AA TANDEM REPEATS OF G-G-A-V-P.
 CC FT REPEAT 266 270 1.
 CC FT REPEAT 271 275 2.
 CC FT REPEAT 276 280 3.
 CC FT REPEAT 281 285 4 (APPROXIMATE).
 CC FT DOMAIN 575 603 9 X 3 AA APPROXIMATE REPEATS OF P-Q-P.
 CC FT SEQUENCE 922 AA; 95178 MW; 3DF7BC58D4712478 CRC64;
 CC -----
 CC RESULT 10
 CC ID PERT_BORPA STANDARD; PRT; 922 AA.
 CC AC P24328;
 CC DT 01-MAR-1992 (Rel. 21, Created)
 CC DT 01-MAR-1992 (Rel. 21, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE Pertactin precursor (outer membrane protein P.70) (P.95).
 CC GN PRN.
 CC OS Bordetella parapertussis.
 CC CC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 CC CC Alcaligenaceae; Bordetella.
 CC CC NCBI_TaxID=519;
 CC CC [1]
 CC CC SEQUENCE FROM N.A.
 CC CC STRAIN=CN2591;
 CC CC MEDLINE=91251771; PubMed=2041476;
 CC CC Li L.J., Dougan G., Novotny P., Charles I.G.;
 CC CC "p.70 pertactin, an outer-membrane protein from Bordetella
 CC CC parapertussis: cloning, nucleotide sequence and surface expression in
 CC CC Escherichia coli.";
 CC CC Mol. Microbiol. 5:409-417(1991).
 CC CC -1- FUNCTION: AGGLUTININ THAT BINDS TO EUKARYOTIC CELLS; A PROCESS
 CC CC MEDIATED BY THE R-G-D SEQUENCE. PERTACTIN MAY HAVE A ROLE IN
 CC CC BACTERIAL ADHESION, AND THUS PLAY A ROLE IN VIRULENCE.
 CC CC -1- SUBUNIT: Monomer.
 CC CC -1- SUBCELLULAR LOCATION: Outer membrane.
 CC CC -1- DISEASE: MAY CONTRIBUTE TO THE DISEASE STATE OF WHOOPING COUGH.
 CC CC -1- MISCELLANEOUS: SYNTHESIZED ONLY IN THE PRESENCE OF LOW MG(2+)
 CC CC CONCENTRATIONS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC DR EMBL; X54547; CA38419.1; -;
 CC DR EMBL; A26124; CA01786.1; -;
 CC DR EMBL; A19182; CA01454.1; -;
 CC DR PIR; S15204; S15204.
 CC DR InterPro; IPR006315; Autotransport.
 CC DR InterPro; IPR005546; Autotransporter.
 CC DR InterPro; IPR004899; Pertactin.
 CC DR InterPro; IPR003991; Pertactin_C.
 CC DR InterPro; IPR003992; Pertactin_N.
 CC Pfam; PF03797; Autotransporter; 1.
 CC Pfam; PF03212; Pertactin; 1.
 CC PRINTS; PRO1482; PERTACTIN.
 CC DR PRINTS; PRO1484; PERTACTINFAMILY.
 CC DR TIGRFAMs; TIGR01414; autocrans Bar1; 1.
 CC KW Outer membrane; Signal; Virulence; Repeat.
 CC FT SIGNAL 1 34 POTENTIAL.
 CC FT CHAIN 35 922 P. 95.
 CC FT CHAIN 35 647 POTENTIAL.
 CC FT PROPEP 648 922 POTENTIAL.
 CC FT SITE 260 922 CELL ATTACHMENT SITE (INVOLVED IN
 CC ADHESION TO VARIOUS EUKARYOTIC CELL
 CC LINES).
 CC FT DOMAIN 266 290 4 X 5 AA TANDEM REPEATS OF G-G-A-V-P.
 CC FT REPEAT 266 270 1.
 CC FT REPEAT 271 275 2.
 CC FT REPEAT 276 280 3.
 CC FT REPEAT 281 285 4 (APPROXIMATE).
 CC FT DOMAIN 575 603 9 X 3 AA APPROXIMATE REPEATS OF P-Q-P.
 CC FT SEQUENCE 922 AA; 95178 MW; 3DF7BC58D4712478 CRC64;

ID	MYCTU	STANDARD;	PRT;	900 AA.
AC	IF2_MYCTU			
DT	IF2_MYCTU			
DT	30-MAY-2000	(Rel. 39, Created)		
DT	15-SEP-2003	(Rel. 39, Last sequence update)		
DT	15-SEP-2003	(Rel. 42, Last annotation update)		
DE	Translation initiation factor IF-2.			
GN	INFB OR RV2839C OR MTCV1687.03.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1773;			
RA	[1]			
RA	SEQUENCE FROM N.A.			
RC	STRAIN=H37Rv;			
RC	MEDLINE=9825987; PubMed=9634230;			
RA	Cole S.T., Broese R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,			
RA	Bacchok R., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R., Devlin K., Feltham T., Genies S., Hamlin N., Holroyd S.,			
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy J.,			
RA	Oliver S., Osborne J., Quail M.A., Rajandream A., Rogers J.,			
RA	Rutter S., Seeger K., Skelton S., Squares R.,			
RA	Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			
RT	complete genome sequence."			
RL	Nature 393:537-544(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CDC 1551 / Oshkosh;			
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,			
RA	Petersen J.F., Deboy R., Dodson R., Gilm M.L., Hatt D., Hickey E.,			
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,			
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,			
RA	Bissh W.;			
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and			
RT	laboratory strains."			
RL	Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.			
CC	-1- FUNCTION: One of the essential components for the initiation of			
CC	protein synthesis. Protects formylmethionyl-tRNA from spontaneous			
CC	hydrolysis and promotes its binding to the 30S ribosomal subunits.			
CC	Also involved in the hydrolysis of GTP during the formation of the			
CC	70S ribosomal complex (By similarity).			
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic.			
CC	-1- SIMILARITY: BELONGS TO THE IF-2 FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; Z81331; CAB03670.1; --			
DR	EMBL; AB007115; AA47231.1; --			
DR	PIR; B70694; B70694.			
DR	HSSP; P13551; IDAR.			
DR	TIGR; MT2905; --			
DR	TubercuList; RV2839C; --			
DR	HMAP; MF 00100; -- 1.			
DR	InterPro: IPR000795; EF_Gnpbind.			
DR	InterPro: IPR004161; EFTU_D2.			
DR	InterPro: IPR000178; IF2.			
DR	InterPro: IPR006847; IF2_N.			
DR	InterPro: IPR001806; Ras_transfmg.			
DR	InterPro: IPR005225; Small_GTP.			
DR	Pfam; PF00009; GTP_EFTU; 1_			
DR	Pfam; PF03144; GTP_EFTU_D2; 2.			
DR	Pfam; PF04760; IF2_N_2.			
DR	PRINTS; PR00315; ELONGATINFCT.			
DR	PRINTS; PR00449; RASTRNSFRMNG.			
DR	ProDom; PD16100; IF2; 1.			

DR TIGRFBMS; TIGR00487; IF-2; 1.
 DR TIGRFBMS; TIGR00231; small GTP, 1.
 DR PROSITE; PS01176; IF2; FALSE NEG.
 KW Initiation factor; Protein biosynthesis; GTP-binding;
 KW Complete proteome.
 FT DOMAIN 399 551 G-DOMAIN
 FT NP_BIND 405 412 GTP (BY SIMILARITY).
 FT NP_BIND 455 459 GTP (BY SIMILARITY).
 FT NP_BIND 509 512 GTP (BY SIMILARITY).
 FT DOMAIN 250 255 POLY-GLY.
 SQ SEQUENCE 900 AA; 94040 MW; F1B2F3C8A86952C2 CRC64;

Query Match 10.5%; Score 116; DB 1; Length 900;
 Best Local Similarity 29.0%; Pred. No. 0.49; Mismatches 50; Indels 62; Gaps 10;
 Matches 49; Conservative 8;

QY 34 PEP-----TKKPSNDIYKPKPPYPPQ-----ENPDSGNIYRPPKPPQ---- 74
 DB 116 PEPNPAVPPSPQAPRPGMAFGARPGVPPKGRTRVGNPPSSAQPADRP1PPEPARPP 175
 QY 75 -----PQPGNSGNS-----GGYFNDVDRDGRYPARP-RRPP-----PAGG 109
 DB 176 GTARPGVPPRPGASPGSMPPRPGAVG-----GAPPRPGAPRPGAPRPGAPGASDAGG 229
 QY 110 G---GGGYS---YGNSDNTHRGGRYRPSRYGNTYGGDHSTYGNPBG 152
 DB 230 GNYRGGGVGAPGTGFRGRPGGGGGGPPGQR-----GGAAGAFRRPG 272

RESULT 13

ASPL_HELAN STANDARD; PRT; 161 AA.

AC P22357;
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE Anterior-specific protein SP18 precursor (Fragment).
 OS Helianthus annuus (Common sunflower).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; campanulids; Asterales; Asteraceae; Asteroideae;
 OC Heliantheae; Helianthus.
 OC NCBI_TaxID=4232;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. HA401B / Caryill; TISSUE=Anterior;
 RA MEDLINE=9138702; PubMed=2102380;
 RX Domon C., Evrard J.-L., Herdenger F., Pillay D.T.N., Steinmetz A.;
 RT "Nucleotide sequence of two anterior-specific cDNAs from sunflower
 (Helianthus annuus L.).";
 RL Plant Mol. Biol. 15:643-646(1990).
 CC -1- FUNCTION: ANTERIOR-SPECIFIC CELL WALL PROTEIN WHICH COULD CONTRIBUTE
 TO THE CELL WALL ARCHITECTURE OF EPIDERMAL ANTERIOR CELLS VIA
 INTERMOLECULAR DISULFIDE BRIDGES.
 CC -1- TISSUE SPECIFICITY: EPIDERMAL ANTERIOR CELLS.
 CC -1- DEVELOPMENTAL STAGE: LATE DEVELOPMENTAL STAGES.
 CC -1- SIMILARITY: BELONGS TO THE PLANT DEFENSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; X53375; CAA37455.1; -
 DR PIR; S12246; S12246.
 DR HSSP; P30231; 1AY7.
 DR InterPro; IPR002118; Gamma-thionin.
 DR InterPro; IPR003614; Knoc1.
 DR Pfam; PF00304; Gamma-thionin; 1.
 DR ProDom; PD002594; G_Purichionin; 1.

DR SMART; SM00505; Knoc1; 1.
 DR PROSITE; PS00940; GAMMA_THIONIN; 1.
 KW Signal; Cell wall.
 FT SIGNAL 1 8
 FT NON_TER 1 1
 FT CHAIN 9 161 ANTERIOR-SPECIFIC PROTEIN SP18.
 FT DOMAIN 9 65 GAMMA-THIONIN LIKE DOMAIN.
 FT NP_BIND 70 161 PROLINE DOMAIN.
 FT DISULFID 18 65 BY SIMILARITY.
 FT DISULFID 29 50 BY SIMILARITY.
 FT DISULFID 35 59 BY SIMILARITY.
 FT DISULFID 39 61 BY SIMILARITY.
 SQ SEQUENCE 161 AA; 15363 MW; 27A9CF4633AD02B CRC64;

Query Match 10.4%; Score 115.5; DB 1; Length 161;
 Best Local Similarity 32.2%; Pred. No. 0.1;
 Matches 39; Conservative 2; Mismatches 37; Indels 43; Gaps 7;

QY 33 DPEPTKKNNSIYKPKPPYTP--QENPDSGNIYRPPKPPQPPQPGNSGSGYFNDV 90
 DB 64 DCDPQKNPG-----PPGAPGTGTPPAPPGKGBDAPHPPTPSP-PGDDGSG----- 112
 QY 91 DRDGRYPRRPPRPPAGG-----GGGGYSYGNSDNTHRGGRYRPSRYGNTYGG 141
 DB 113 -----PAPPGGSGPPAGGDDGGGAPPPAGD---GGGAPPA-----GG 151
 QY 142 D 142
 DB 152 D 152

RESULT 14

PRIO_BOVIN STANDARD; PRT; 264 AA.

AC P10279;
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 01-NOV-1991 (Rel. 20, Last annotation update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Major prion protein 1 precursor (Prp) (Major scrapie-associated fibril
 protein 1).
 DE PRNP OR PRP.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
 OC Mammalia; Butheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
 OC Bovidae; Bovinae; Bos.
 OC NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Holstein-Friesian;
 RA MEDLINE=9116314; PubMed=1671225;
 RX Goldmann W., Hunter N., Martin T., Dawson M., Hope J.;
 RT "Different forms of the bovine Prp gene have five or six copies of a
 short, G-C-rich element within the protein-coding exon";
 RL J. Gen. Virol. 72:201-204(1991).
 CC [2]
 CC SEQUENCE FROM N.A.
 CC TISSUE=Brain;
 CC MEDLINE=93118243; PubMed=1362024;
 RA Yoshimoto J., Iinuma T., Ishiguro N., Horiuchi M., Imamura M.,
 RA Shinagawa M.;
 RT "Comparative sequence analysis and expression of bovine Prp gene in
 RT mouse U-929 cells";
 RL Virus Genes 6:343-356(1992).
 CC [3]
 CC SEQUENCE FROM N.A.
 CC MEDLINE=93179783; PubMed=8440932;
 RX Prusiner S.B., Puzi M., Scott M., Serban D., Serban H.,
 RA Taraboulos A., Gabriel J.M., Wells G.A., Mueselich J.W., Bradley R.;
 RT "Immunologic and molecular biologic studies of prion proteins in
 RT bovine spongiform encephalopathy";
 RL J. Infect. Dis. 167:602-613(1993).
 CC [4]
 CC SEQUENCE FROM N.A.

CC STRAIN=Holestein-Friesian; TISSUE=Brain;
RA Horiuchi M.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN (5)
RP SEQUENCE FROM N.A.
RC STRAIN=Jersey;
RX MEDLINE=21422903; PubMed=11531705;
RA Hille D., Comincini S., Schlaepfer J., Dolf G., Ferretti L.,
RA Williams J.L.;
RT "Complete genomic sequence of the bovine prion gene (PRNP) and
RT polymorphism in its promoter region.";
RL Anim. Genet. 32:231-232(2001).
RN (6)
RP SEQUENCE FROM N.A.
RC STRAIN=Korean;
RX Yoo H.S., Kang S.G., Choi I.S., Kang S.K., Hwang W.S.;
RA "Nucleotide sequence of PrP cDNA in Korean cattle.";
RT Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
RL (7)
RP SEQUENCE OF 1-15 FROM N.A.
RA Tanaka M., Inoue S., Ikeda T., Horiuchi M., Ishiguro N., Shinagawa M.;
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
RN (8)
RP SEQUENCE OF 25-36.
RX MEDLINE=89057122; PubMed=2904126;
RA Hope J., Reekie L.J.D., Hunter N., Multhaup G., Beyreuther K.,
RA White H., Scott A.C., Stack M.J., Dawson M., Wells G.A.;
RT "Fibrils from brains of cows with new cattle disease contain scrapie-
RT associated protein.";
RL Nature 336:390-392(1998).
RN (9)
RP STRUCTURE BY NMR OF 132-241.
RX MEDLINE=20359707; PubMed=10899999;
RA Lopez Garcia F., Zahn R., Riek R., Wuehrich K.;
RT "NMR structure of the bovine prion protein.";
RT Proc. Natl. Acad. Sci. U.S.A. 97:8334-8339(2000)
CC -1- FUNCTION: THE FUNCTION OF PRP IS NOT KNOWN. PRP IS ENCODED IN THE
CC HOST GENOME AND IS EXPRESSED BOTH IN NORMAL AND INFECTED CELLS.
CC -1- SUBUNIT: PRP HAS A TENDENCY TO AGGREGATE YIELDING POLYMERS CALLED
CC "RODS".
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- DISEASE: FOUND IN HIGH QUANTITY IN THE BRAIN OF HUMANS AND ANIMALS
CC INFECTED WITH DEGENERATIVE NEUROLOGICAL DISEASES SUCH AS KURU,
CC CRETZFELDT-JAKOB DISEASE (CJD), GERSTMANN-STRAUSSLER SYNDROME
CC (GSS), SCRAPIE, BOVINE SPONGIFORM ENCEPHALOPATHY (BSE),
CC TRANSMISSIBLE MINK ENCEPHALOPATHY (TME), ETC.
CC -1- SIMILARITY: BELONGS TO THE PRION FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; X55882; CAA39368.1; -
DR EMBL; D10612; BAA01467.1; -
DR EMBL; D10613; BAA01468.1; -
DR EMBL; S55629; AAB25514.1; -
DR EMBL; AJ298878; CAC37367.1; -
DR EMBL; AF517842; AAM66709.1; -
DR EMBL; D26151; BAA05138.1; -
DR PIR; A54330; A54330.
DR PDB; 1DWY; 26-FEB-02.
DR PDB; 1DWZ; 26-FEB-02.
DR PDB; 1DX0; 26-FEB-02.
DR PDB; 1DX1; 26-FEB-02.
DR InterPro; IPR000817; Prion.
DR Pfam; PF00377; Prion; 1.
DR Pfam; PF03991; Prion octapep. 6.
DR PRINTS; PRO0341; PRION.

DR SMART; SM00157; PRP; 1.
DR PROSITE; PS00291; PRION 1; 1.
DR PROSITE; PS00706; PRION 2; 1.
KW Prion; Brain; Glycoprotein; GPI-anchor; Repeat; Signal; Polymorphism;
KW 3D-structure.
FT SIGNAL 1 24
FT CHAIN 25 264 MAJOR PRION PROTEIN 1.
FT CARBOHYD 192 192 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 208 208 N-LINKED (GLCNAC. . .) (PROBABLE).
FT DISULFID 190 225
FT DOMAIN 54 103 6 X 8 AA TANDEM REPEATS OF P-H-G-G-W-G-G-
FT
FT REPEAT 54 62 0.
FT REPEAT 63 70 2.
FT REPEAT 71 78 3.
FT REPEAT 79 86 4.
FT REPEAT 87 94 5.
FT REPEAT 95 103 6.
FT REPEAT 71 78 MISSING (IN ALLELE 2).
FT VARIANT 218 218 E -> K (IN REF. 2 AND 4).
FT CONFLICT 218 218
FT HELIX 136 138
FT STRAND 140 141
FT STRAND 140 141
FT TURN 155 162
FT HELIX 163 164
FT HELIX 165 167
FT STRAND 173 174
FT HELIX 184 203
FT TURN 204 206
FT HELIX 211 237
SQ SEQUENCE 264 AA; 28614 MW; D6D21403B16A231 CRC64;
Query Match 10.4%; Score 115; DB 1; Length 264;
Best Local Similarity 32.9%; Pred. NO. 0.18;
Matches 28; Conservative 9; Mismatches 32; Indels 16; Gaps 3;
QY 72 RQPPGSGNSGGYFNDVDDGGRPRPRP-----RPPAGGGGGGSGSYGSDNTHRG 127
DB 25 KRPFGGGGWNMG-----SRYPGGSGGNRYPFGGGGSGWGPFGGG 75
QY 128 GYFNSRYGTYGGDHSYGNPBG 152
DB 76 WQGP---HGGMGQPHGGMGQPHG 97
RESULT 15
IP2 MYCLE STANDARD; PRT; 924 AA.
AC 092519;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Translation initiation factor IF-2.
GN INFB OR M1556 OR MCB596.14.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OC NCBI_TaxID=1769;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltham T., Fraser A., Hamlin N.,
RA Holroyd S., Hornby T., Jagerle K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds W., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -1- FUNCTION: One of the essential components for the initiation of

protein synthesis. Protects formylmethionyl-tRNA from spontaneous hydrolysis and promotes its binding to the 30S ribosomal subunits. Also involved in the hydrolysis of GTP during the formation of the 70S ribosomal complex (by similarity).

-1 SUBCELLULAR LOCATION: Cytoplasmic.

-1 SIMILARITY: BELONGS TO THE IF-2 FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR	EMBL; AL0354472; CAB36570.1; -.
DR	EMBL; AL583922; CAC30507.1; -.
DR	PIR; P87103; P87103.
DR	HSSP; P1351; 1ELO.
DR	Leptoma; ML1556; -.
DR	HAMAP; MF_00100; -; 1.
DR	InterPro; IPR000795; BP_GTPbind.
DR	InterPro; IPR004161; EFTU_D2.
DR	InterPro; IPR000178; IP2.
DR	InterPro; IPR006847; IP2_N.
DR	InterPro; IPR05325; Small_GTP.
DR	Pfam; PF00009; GTP_EFTU; 1.
DR	Pfam; PF03144; GTP_EFTU_D2; 2.
DR	Pfam; PF04760; IP2_N; 2.
DR	PRINTS; PR00315; ELONGATNPCT.
DR	ProDom; T186100; IP2; 1.
DR	TIGRFAMS; TIGR00487; IP-2; 1.
DR	TIGRFAMS; TIGR00231; small_GTP; 1.
DR	PROSITE; PS01176; IP2; FALSE NEG.
KW	Initiation factor; Protein biosynthesis; GTP-binding;
KW	Complete proteome.
FT	DOMAIN 423 575 G-DOMAIN.
FT	NP_BIND 429 436 GTP (BY SIMILARITY).
FT	NP_BIND 479 483 GTP (BY SIMILARITY).
FT	NP_BIND 533 536 GTP (BY SIMILARITY).
FT	DOMAIN 216 219 POLY-GLY.
FT	DOMAIN 271 279 POLY-GLY.
SO	SEQUENCE 924 AA; 96650 MW; 34605A1069001224 CRC64;

Query Match	10.4%	Score 115	DB 1	Length 924
Best Local Similarity	31.0%	Pred. No. 0.59		
Matches 44	Conservative 3	Mismatches 49	Indels 46	Gaps 8

```

Oy  46 KKPKEPPYYPQ--PKNPDSGNTLYPPKPKP-----QPQKGSNGSGYENVD  91
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  166 PAPKGGKGPPIRVYNNPFSSAQSVARPIPPPPAPRPSASPSMSPPRGAVGGG----P  220
Oy  92 RDDGRYPKRP---RPRPACG-----GGGGYS-----YGNSDNTHGR-----GGYR  130
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  221 RPRRTGVPPPGGGRGAPVGGGSDAGGNGTGGVGVALLPGGGSGGFTGRPGGGHGGGR  280
Oy  131 PMSRYGNTYGGEDHHSITYGNPEG  152
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  281 PGQR-----GSAAGAFGRPGG  296
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: February 18, 2004, 19:05:43
Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2004, 19:02:25 ; Search time 35 seconds
(without alignment)
1437.722 Million cell updates/sec

Title: US-09-997-701-1

Perfect score: 1110

Sequence: 1 MESMGWGLCLAFCLFAMHAR.....SYFKLNRRNCRTHPEENV 195

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteria:*
17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	181	16.3	175	11	Q8VCN6
2	167.5	15.1	174	11	Q8BLN1
3	165.5	14.9	157	6	Q46670
4	164.5	14.8	173	6	Q02788
5	149.5	13.5	213	4	Q8TC21
6	149.5	13.5	419	16	Q8NVK9
7	149	13.4	262	4	Q8TC22
8	147	13.2	134	11	Q8C1L8
9	146	13.2	235	13	Q8QFW9
10	146	13.2	415	16	Q8QFW9
11	143	12.9	354	10	Q8FYEA
12	142	12.8	1046	10	Q8H093
13	141	12.7	190	4	Q8TC20
14	141	12.7	521	16	Q8VKT1
15	140	12.6	461	13	Q8AX89
16	139	12.5	527	16	P71590

17	138.5	12.5	246	11	Q8R1R5	Q8R1R5 rattus norv
18	138.5	12.5	374	11	Q9JUL7	Q9JUL7 rattus norv
19	138.5	12.5	300	10	Q07373	Q07373 chlamydomon
20	135	12.2	381	10	Q9FKA5	Q9FKA5 arabidopsis
21	135	12.2	381	10	Q8H7G9	Q8H7G9 arabidopsis
22	135	12.2	594	5	Q9VEP4	Q9VEP4 drosophila
23	134.5	12.1	108	5	Q9GTX8	Q9GTX8 fasciola he
24	134.5	12.1	880	16	Q8F7K1	Q8F7K1 leptospira
25	133	12.0	735	3	Q9UVQ3	Q9UVQ3 podospira a
26	132.5	11.9	623	11	Q8K2U8	Q8K2U8 mus musculu
27	132	11.9	460	5	Q8BIF0	Q8BIF0 mus musculu
28	132	11.9	237	11	Q9VFE6	Q9VFE6 drosophila
29	132	11.9	595	10	Q8S1W9	Q8S1W9 oryza sativ
30	130.5	11.8	164	10	Q8L1U1	Q8L1U1 saccharum h
31	130.5	11.8	2344	5	Q9N3Y8	Q9N3Y8 caenorhabdi
32	130	11.7	463	5	Q93535	Q93535 caenorhabdi
33	127.5	11.5	247	10	Q8H5X2	Q8H5X2 oryza sativ
34	127.5	11.5	247	10	Q8H5W8	Q8H5W8 oryza sativ
35	127.5	11.5	637	4	Q8BSV4	Q8BSV4 homo sapien
36	127.5	11.5	872	11	Q55000	Q55000 rattus norv
37	127	11.4	497	5	Q27512	Q27512 caenorhabdi
38	125.5	11.3	358	10	Q8W2E5	Q8W2E5 oryza sativ
39	125.5	11.3	374	2	Q33899	Q33899 streptococc
40	125.5	11.3	513	10	Q91V82	Q91V82 arabidopsis
41	125.5	11.3	513	10	Q8W4P1	Q8W4P1 arabidopsis
42	125.5	11.3	525	11	Q99P86	Q99P86 mus musculu
43	125.5	11.3	525	11	Q99P85	Q99P85 mus musculu
44	125.5	11.3	525	11	Q9ESB3	Q9ESB3 mus musculu
45	125.5	11.3	1379	5	Q8SWX1	Q8SWX1 drosophila

ALIGNMENTS

RESULT 1

Q8VCN6 PRELIMINARY: PRT: 175 AA.
AC Q8VCN6 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
BT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Similar to RIKEN CDNA 2410026K10 gene.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RA Strusberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC019482; AAH19482.1;
SQ SEQUENCE 175 AA; 16782 MW; A309B6E2494FBE1 CRC64;

Query Match 16.3%; Score 181; DB 11; Length 175;
Best Local Similarity 30.0%; Pred. No. 2.3e-08;
Matches 60; Conservative 17; Mismatches 55; Indels 68; Gaps 10;

QY	10	LAFLCFLMHARGQRFDLADLADPEPTKKNPSDIYPPKPPYYPQENPDSG-----	62
DB	16	LALGGAARGAASDPFNIGDALBDP-----NMKTPPKA---PTPKKPSGFLLEDA	64
QY	63	-----GNTPPKPPKPPQPPQPNSGSGYFNDVDRDGRYPKPPKPPKPPAGG	110
DB	65	PGGGGGGAGKPPGN---RPPDPKPK---PPPHDSCG---ISDSDLADA-----	108
QY	111	GGGYSSVYNSD---NTHGRGVRPNRSGYNTYGGDHSTYGNPEGMMVAKIVSPIYVVV	168
DB	109	GGAGRRSGDBGGHGGAGAPR-----GTPQG-----LVPGVAAVVA	148
QY	169	TLGGAASYFKLNRRNCR	188
DB	149	AVAGVSSFFVAYQRRRLCFR	168

RESULT 2

OB0BLN1 PRELIMINARY; PRT; 174 AA.
 AC OB0BLN1;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Hypothetical glycine-rich region containing protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 NC NCB1_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CS7BL/6J; TISSUE=Cortex;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR EMBL; AK044051; BAC1754.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 174 AA; 16719 MW; 580017FD53A4B18 CRC64;

Query Match 15.1%; Score 167.5; DB 11; Length 174;
 Best Local Similarity 31.5%; Pred. No. 3.5e-07;
 Matches 57; Conservative 20; Mismatches 73; Indels 31; Gaps 10;

QY 10 LAFICPLMHARGQDFDLADLDDEPTKPNSDIYKPKPPYY----- 69
 DB 16 LALLGGAARGAASDPLGDLLEP-----NMKPTPKA---PTPKSGGFDL----- 60
 QY 70 KPRPOPGNSGSGYFNVDYRDDGRYPPRRPP-AGGGGGYSGNSDTHGRGG 128
 DB 61 ---BDALPGGGGGAG-----EKPNR--PQPDKPPPHDSGDISDGLAD--AAQGG 109
 QY 129 YRPSRYGNTYGGDHSTYV--NPEGNNVAKIVSIYVVTLLGAAAYPKLNRRNCF 187
 DB 110 --GAGRGSGDEGGHGAAGABEGTPOG-LVPGVAVAVPPVAGAVSSFAVYGRRLCF 166
 QY 188 R 188
 DB 167 R 167

RESULT 3

OB0BLN1 PRELIMINARY; PRT; 157 AA.
 AC 046670;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE CD99 type II-COS7 (Fragment).
 OS Cercopithecus aethiops (Green monkey) (Grivet).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Cercopithecus.
 NC NCB1_TaxID=9534;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Park S.H.;
 RT "A monkey homolog of an alternative spliced form of human CD99."
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U82166; AAB93833.1; -.
 FT NON TER 1
 SQ SEQUENCE 157 AA; 15679 MW; C34DB239B4D55E20 CRC64;

Query Match 14.9%; Score 165.5; DB 6; Length 157;
 Best Local Similarity 27.1%; Pred. No. 4.6e-07;
 Matches 55; Conservative 19; Mismatches 58; Indels 71; Gaps 8;

QY 10 LAFICF-----LMHARGQDFDLADLDDEPTKPNSDIYKPKPPYY----- 53
 DB 1 LALLFGLGALVAAADDGFDLSALPEKE-DKKPTA-----TPKKSAGDDPLGAAVVD 55
 QY 54 -----PQENPDGSGNIYPRKPPPOPGNSGSGYFNVDYRDDGRYPPRRPPAG 108
 DB 56 GGNDDEPPPP-----PKKPNPNPQAGSSGS-FSDADLDAG-----VSG 95
 QY 109 GGGGYSYSGNSDTHGRGGRYPNRYGNTYGGDHSTYGNPBNVAKIVSIYVWV 168
 DB 96 GEGKGGSDGGSGPRKGEADAPG-----VIGIVGAVV 130
 QY 169 TLGAAAYPKLNRRNCFRTH 191
 DB 131 AVAGAISSFLAYOKKLCFRKAN 153

RESULT 4

OB02788 PRELIMINARY; PRT; 173 AA.
 AC 002788;
 DT 01-JUL-1997 (TREMBlrel. 04, Created)
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE CD99-cos7 (Fragment).
 OS Cercopithecus aethiops (Green monkey) (Grivet).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Cercopithecus.
 NC NCB1_TaxID=9534;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Park S.H., Hahn J.-H., Kim M.K., Sohn H.W., Choi E.Y., Kim S.H.;
 RT "A monkey homolog of human CD99."
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U82165; AAB58504.1; -.
 FT NON TER 1
 SQ SEQUENCE 173 AA; 17299 MW; B9667ABED16287B8 CRC64;

Query Match 14.8%; Score 164.5; DB 6; Length 173;
 Best Local Similarity 27.2%; Pred. No. 6.3e-07;
 Matches 55; Conservative 18; Mismatches 58; Indels 71; Gaps 8;

QY 10 LAFICF-----LMHARGQDFDLADLDDEPTKPNSDIYKPKPPYY----- 53
 DB 1 LALLFGLGALVAAADDGFDLSALPEKE-DKKPTA-----TPKKSAGDDPLGAAVVD 55
 QY 54 -----PQENPDGSGNIYPRKPPPOPGNSGSGYFNVDYRDDGRYPPRRPPAG 108
 DB 56 GGNDDEPPPP-----PKKPNPNPQAGSSGS-FSDADLDAG-----VSG 95
 QY 109 GGGGYSYSGNSDTHGRGGRYPNRYGNTYGGDHSTYGNPBNVAKIVSIYVWV 168
 DB 96 GEGKGGSDGGSGPRKGEADAPG-----VIGIVGAVV 130
 QY 169 TLGAAAYPKLNRRNCFRTH 190
 DB 131 AVAGAISSFLAYOKKLCFRKAN 152

RESULT 5

OB02C21 PRELIMINARY; PRT; 213 AA.
 AC 082C21;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE MIC2L1 isoform E3-84.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

Ox	NCl TaxID=9606;
Rn	[1]
Rp	SEQUENCE FROM N.A.
Rc	TISUB=Skeletal muscle;
Rt	Suh Y.H., Shin Y.K., Park S.H.;
Rl	"MIC2L1, a novel human paralog of CD99 (MIC2), E3-E4 isoform.";
Rd	Submitted (FEB-2002) to the EMBL/genbank/DDOJ databases.
Dk	EMBL; AY078166; AAL86618.1; ..
So	SEQUENCE 213 AA; 22833 MW; F90595DDAA9BA7D5 CRC64;
Qy	Query Match 13.5%; Score 149.5; DB 4; Length 213; Best Local Similarity 27.1%; Pred. No. 1.7e-05;
Matches	55; Conservative 29; Mismatches 80; Indels 39; Gaps 9
Db	1 MESWGG--LPCLAFGLFMHARGRDF---LADALDPDEPTKPKNSDIYPKPPEPPYQ 55 : 1 MVAMRSATFLVCIASFSLATLVDRSGSDFDFFNLDAVKETSSVKQRMNVTTLTKRPVTR 60 : Qy 56 PENPDSCGNIPRPKRPRPQPQGNGSNGGYYND-VDRDXGHYPRRPRPRPAGGGGGY 114 : Db 61 -----APANTLGDNFDLADALDENRDDR-----RKFIAGGG-- 95 : Qy 115 SSYGNSDTHRGGRGYRPSRYGN-TYGGDHSTYGNPGNNVAK--IYSPIVSVVVLL 171 : Db 96 PSDXDLEIYVGGEYFKDKGDGKYGSN----DDPSGVVAEPGTIGVASALMALLI 150 : Qy 172 GAASLYFKLNRRNCFRTERPEN 194 : Db 151 GAVSSYSIQOKKFCSIQGLIN 173 :
Result	6
ID	QBNAVX9 PRELIMINARY; PRT; 419 AA. QBNAV9 AC QBNAK9;
DT	01-OCT-2002 (TREMblrel. 22, Created)
DT	01-OCT-2002 (TREMblrel. 22, Last sequence update)
DT	01-MAR-2003 (TREMblrel. 23, Last annotation update)
De	MM1956 protein.
Gn	Staphylococcus aureus (strain MM2).
OS	Bacteria; Firmicutes; Bacillales; Staphylococcus.
Ox	NCl Taxid=196620;
Rn	[1]
Rp	SEQUENCE FROM N.A.
Rc	MEDLINE=22040717; PubMed=12044378;
Rt	Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
Rl	Nagai Y., Ikema N., Asano K., Naito T., Kuroda H., Cui L.,
Rd	Yamamoto K., Hiratacsu K.;
Rt	"Genome and virulence determinants of high virulence community-
Rl	acquired MRSA."
Rl	Lancet 359:1819-1827 (2002). RL EMBL: AF004829; BAB9821.1; -
Rd	InterPro: IPR002965_P-rich_extensin.
Dr	PRINTS; PRO1217; PRICHEXTENSIN.
Kw	Complete proteome.
So	SEQUENCE 419 AA; 46630 MW; E31ED7F6525A1B18 CRC64;
Qy	Query Match 13.5%; Score 149.5; DB 16; Length 419; Best Local Similarity 27.0%; Pred. No. 3.5e-05;
Matches	44; Conservative 17; Mismatches 55; Indels 47; Gaps 6;
Dk	34 PEPTKKNSDIYPKPPEPPYQPEZPDSCGNIY-PKRPPOFGNSGSGGFNDV 90 : Db 126 KPDPKXPDPD-KPKRNPDPKDPDNPKENPPKCDPKCKKNPDPKP-----DP 173 : Qy 91 DRDDGARYPRPRPPP-----AGGGGGSYSGNSDNTHRGGYRPNR 134 : Db 174 DKPKRNPNPKEDPKKPNPNPSPDDQPDGDSNHSGSKKGWTNPNASDGSMQOWPYGN 233 : Qy 135 YGNITVGDDHSTYGNPGSNVAKIVSPIVSVVVTTLGAASY 177 : Db 234 QGNSQ-----NFTGN-----DFVSGRFALANGARKY 260 :

ID	Q8TCZ2	PRELIMINARY;	PRT;	262 AA.
AC	Q8TCZ2;			
DT	01-JUN-2002 (TREMBLrel. 21, Created)			
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)			
DT	01-OCT-2002 (TREMBLrel. 22, Last annotation update)			
DE	MIC2L1 isoform E3'-B4'-E3-B4 (MIC2, like 1).			
OS	Homo sapiens (human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RP	SEQUENCE FROM N.A.			
RA	Sub Y.H., Shin Y.K., Park S.H.;			
RT	"MIC2L1, a novel human paralog of CD99 (MIC2), E3'-B4'-E3-B4			
RL	isoform."			
RL	Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Brain;			
RA	Strausberg R.;			
RL	Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AY078165; AAL86617.1;			
DR	EMBL; BC030536; AAB30536.1;			
DR	Genew; HGNC:18237; MIC2L1.			
SO	SEQUENCE 262 AA; 27986 MW; 3A254961DD32C191 CRC64;			
Query Match	13.4%;	Score 149;	DB 4;	Length 262;
Best Local Similarity	25.4%;	Pred. No. 2.3e-05;		
Matches	60;	Conservative 29;	Mismatches 91;	Indels 56; Gaps 11;
Qy	1 MESWVG--LPCIAVCFMFAHARGGRDPD--LADALDDPEPTKCP--NSDIYPKPKPPY 53			
Db	1 MVAARSAFIVCIASFSLATIVQSGGCDPDTFLBDAVETSSVQKPMHITTTTNNRPQT 60			
Qy	54 POPENPDSG-----NTPPKPPPPQPOGNSGSGYFN 88			
Db	61 RAFAKPPGSGGLDALDADDGGRKPKGIGGEREMNVHTTTTKRPVTTTRAPANTLGDNFD 120			
Qy	89 --DV-----RRDGRYFPRPRPPAGGGGGYSSYNSNTHRGGRGPRASRYGN--TYG 140			
Db	121 LADLDDRRDDDR-----KPLVAGGGG--PSDKOLBIVGGGKPKPKGGKGGRYG 171			
Qy	141 GDHHSITGVNPEGNNVAK--IVSPVSVVVVTLTGAAASYFKLANRRCFRTHEPEN 194			
Db	172 SN-----DDPGSGGVABPGTLAGVASLMAALIGAVSSYSYQKKKCFSTIGQLN 222			
RESULT 8				
ID	Q8CIL8	PRELIMINARY;	PRT;	134 AA.
AC	Q8CIL8;			
DT	01-MAR-2003 (TREMBLrel. 23, Created)			
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)			
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)			
DE	Hypothetical glycoline-rich region containing protein.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J;			
RA	MEDLINE=2354683; PubMed=12466851;			
RA	the PANTOM Consortium,			
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;			
RT	"Analysis of the mouse transcriptome based on functional annotation of			
RT	60,770 full-length cDNAs."			
RL	Nature 420:563-573 (2002).			
RL	EMBL; AK010597; BAC25305.1;			

DR EMBL: AL391716; CAC05499.1; -
 DR EMBL: AY081342; AAL91231.1; -
 DR InterPro: IPR002048; EF-hand.
 DR Pfam: PF00036; ehand; 2.
 DR ProDom: PD000012; EF-hand; 1.
 DR SMART: SM00054; EFh; 2.
 DR PROSITE: PS00018; EF HAND; 2.
 SQ SEQUENCE 354 AA; 37113 MW; F827C8342E258FA2 CRC64;

Query Match 12.9%; Score 143; DB 10; Length 354;
 Best Local Similarity 31.4%; Pred. No. 0.00011;
 Matches 49; Conservative 15; Mismatches 52; Indels 40; Gaps 8;

QY 36 PTKKPSDLYPKPKPPYPP-----ENPDSGANTPRPKPRPQPQPGNSGSGYF 87
 DB 63 PTVSSSYCAPPSAPAYASPDYKNPKPKYGGG--YGAPP-----SGSS----- 108
 QY 88 NDVDRDDGRPPRRPRRPPAGGGG-----GYSYSGNDTHGGRYP-NSRGRNTY 139
 DB 109 -----DYSGYGAQPRPSQPSHGCGYGATPRPHGVSDYG---SYGAPPRPASGHGGY 159
 QY 140 GG-DHHSYTGNEGNNVAKIVSVVVVTLGAA 174
 DB 160 GGYPQASYSQSPFASLIPSGFAPGTDPNTVACFGQA 195

RESULT 12

Q8H093 PRELIMINARY; PRT; 1046 AA.

AC Q8H093; 01-MAR-2003 (TREMblrel. 23, Created)
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE Putative phospholipase.
 GN OSJNBA0096G08.23.
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhacoidae; Oryzae; Oryza.
 OC NCBI_TaxId=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Buell C.R., Yuan Q., Ouyang S., Liu J., Ganabinger K., Kim M.M.,
 RA Overton II L.L., Bera J.J., Taitir T., Krol M.L., Dattah B.B.,
 RA Jin S.S., Koo H., Ziemann V., Hsieh J., Blunt S., Vanaken S.S.,
 RA Utecherback T.T., Feldblum T.V., Yang Q.Q., Haas B.J., Suh B.B.,
 RA Peterson J.J., Quackenbush J., White O., Salzberg S.L., Frazer C.M.;
 RT "Oryza sativa chromosome 10 BAC OSJNBA0096G08 genomic sequence."
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC078894; AAO00720.1; -
 SQ SEQUENCE 1046 AA; 116544 MW; BBA3A8C1985EA577 CRC64;

Query Match 12.8%; Score 142; DB 10; Length 1046;
 Best Local Similarity 25.2%; Pred. No. 0.00045;
 Matches 51; Conservative 18; Mismatches 67; Indels 66; Gaps 9;

QY 23 RDPDLADLDDPEPTKP-----NSDIYKPKPPY-----POPENP-----SG 62
 DB 48 RSFHGYPSPAPQPOQFVAHHSAPLOYPRRPHNAVRRPDRPHSPRYVDYPAHAA 107
 QY 63 GNIYRPRKRP-----OPQNSGNSGGRNDV-----DRDDGRPPRR 102
 DB 108 YPSYSPNSPSPISSSSFHHPBPSPSPASPSYSLADGLANMHVSDRHD--YPPPS 165
 QY 103 PR-----PAGGGGGYSSYSGNSDTHGGRYPNRYGNTYGGDHS-----T 146
 DB 166 PAAYAAASPSYLPSPASPRGGSSHG-----CGGMQVYPRPPAGSGHGGMQVYA 217
 QY 147 YGNEPGNNVAKIVSVVVV 168
 DB 218 YGSPAGSGHSGVSRPSLKVLL 239

RESULT 13

Q8TC20 PRELIMINARY; PRT; 190 AA.

AC Q8TC20; 01-JUN-2002 (TREMblrel. 21, Created)
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Mic2L1 isoform E4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Suh Y.H., Shin Y.K., Park S.H.;
 RT "Mic2L1, a novel human paralog of CD99 (MIC2), E4 isoform."
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY078167; AAL86619.1; -
 SQ SEQUENCE 190 AA; 20300 MW; F21A43C7F6B319F6 CRC64;

Query Match 12.7%; Score 141; DB 4; Length 190;
 Best Local Similarity 23.7%; Pred. No. 8.2e-05;
 Matches 52; Conservative 22; Mismatches 51; Indels 94; Gaps 8;

QY 1 MESWNG--LPCLAFCLMHARG-----ORDPLADLDDPEPTK 38
 DB 1 MVAMRSAPLVCLARSLATLVORSGGDFDNLDAVKETSSVKRNDPLADLDDRN--- 57
 QY 39 KPNSDIYKPKPPYPPPENPDSGNTYPRPKPRPQPQPGNSGSGYFNDVDRDGRYP 98
 DB 58 -----DRDDGR-- 63
 QY 99 PRPRPPAGGGGGYSSYSGNSDTHGGRYPNRYGN-TYGGDHSYTGNEGNNVAK 157
 DB 64 -----RKRIAGGG--PSDKLELDIVGGGEYKPKGKDGKDGYSN-----DDPGSWYAE 111
 QY 158 --IVSPISVVVVTLGMAASYFKLNNRNCFRTHPEPN 194
 DB 112 PGTIAGVASALMALIGAVSSYSISYQKKCFCSIOGGLN 150

RESULT 14

Q8VKT1 PRELIMINARY; PRT; 521 AA.

AC Q8VKT1; 01-MAR-2002 (TREMblrel. 20, Created)
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)
 DE Hypothetical protein MT0023.
 GN MT0023.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OC NCBI_TaxId=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ernoldeva M.D., Salzberg S.L.,
 RA Delcher A., Utecherback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF006916; AAK44245.1; -
 DR TIGR; MT0023; -
 DR InterPro; IPR000253; FHA.
 DR Pfam; PF00498; FHA; 1.
 DR SMART; SM00240; FHA; 1.
 DR PROSITE; PS5006; FHA_DOMAIN; 1.
 KW Hypothetical protein.

SEQ SEQUENCE 521 AA; 56149 MW; 6903AB1D396A9C36 CRC64;

Query Match 12.7%; Score 141; DB 16; Length 521;

Best Local Similarity 31.2%; Pred. No. 0.00025; Matches 50; Conservative 12; Mismatches 66; Indels 32; Gaps 10;

QY 18 HARGORDFDLADLDDEPTKPKNSDIYPR-----PKPPYYQPEPNPDGSGNIYPR-----68
 DB 179 YARPQED---PRGSPDQGGSDPRGYPPEYGGYPPQPGYPRPRHPDQD--YPEQIGYR 233
 QY 69 -----PKPRPOP-OPGNSGSGYFENDVDKRYPPRPPRPPAG-GGGGYSYGNSD 121
 DB 234 DGGYPPQRGYPDQGGYQDQGRG---PDQGGGYPPPEQRPPVSPPAAGYGAFCYDQ 290
 QY 122 NTHGSGYRPNRSRYGNT-YGGDHSYTG-----NPEGNMV 155
 DB 291 GYRSGGYGSPGSGQPGYGG--YGEYGRGPARHEGSYV 328

RESULT 15

OBAX89 PRELIMINARY; PRT; 461 AA.

AC OBAX89; 01-MAR-2003 (TRENBLREL. 23, Created)
 DT 01-MAR-2003 (TRENBLREL. 23, Last sequence update)
 DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
 DE Prion-like protein.
 GN PRP-LIKE.
 OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Takifugu.
 CX NCBI_TaxID=31033;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rivera-Milla E.L., Malaga-Trillo E.G., Stuermer C.A.O.;
 RT "An evolutionary basis for scrapie disease: identification of a fish
 prion."
 RL Trends Genet. 0:0-0(2003).
 DR EMBL; AF531159; AAN8028.1;
 SO SEQUENCE 461 AA; 48913 MW; 619C1E5442C10C36 CRC64;

Query Match 12.6%; Score 140; DB 13; Length 461;

Best Local Similarity 34.1%; Pred. No. 0.00027; Matches 44; Conservative 6; Mismatches 53; Indels 26; Gaps 6;

QY 38 KKPSNDIYPRKPPYQPEPNPDGSGNIYPRKPPQPGPNSGNSGY-----FNDVD 91
 DB 35 KKTSSSTWGSRTSPKPSGSGQPNRNSNYPGSGSYPOGTQS-NPGYPRQNPASNPVG 93
 QY 92 RDDGRYPPRPP-----RPPAGG-----GGYSY-----GNSDTHGRGYRPN--- 132
 DB 94 GSPNQYPRNTPGYPPNPGAGGYPNQNPAGGYPNQNPARGNYPNQYPPAGSNNPQYP 153
 QY 133 SRYGNTYGG 141
 DB 154 GRAGTNGG 162

Search completed: February 18, 2004, 19:06:31
 Job time : 37 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2004, 18:56:15 ; Search time 40 Seconds

(without alignments)
773.792 Million cell updates/sec

Title: US-09-997-701-1

Perfect score: 1110

Sequence: 1 MBSWGLPCIAFLCFLMAR.....SYFLANRNCFTHEPENY 195

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1999.DAT:*
21: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA2000.DAT:*
22: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA2001.DAT:*
23: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA2002.DAT:*
24: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1110	100.0	195	21	AAW90961 Human CSGP-1 prote
2	1110	100.0	195	23	ABG73109 Human cell surface
3	885	79.8	168	22	AAW23909 Human EST encoded
4	502	45.2	96	22	ABG00680 Novel human diagno
5	491	44.2	91	21	AAW00621 Human secreted pro
6	465	41.9	345	22	ABG27670 Novel human diagno
7	427	38.5	91	22	ABG27669 Novel human diagno
8	371.5	33.5	73	22	ABG00689 Novel human diagno
9	308	27.7	62	22	ABG03178 Novel human diagno

10	176.5	15.9	169	21	AAW19128 Polypeptide isolat
11	172.5	15.5	165	23	ABW72314 Rat protein isolat
12	164.5	14.8	195	21	AAW58379 Lung cancer associ
13	164.5	14.8	195	24	ABG04013 Human expressed pr
14	162.5	14.6	265	22	ABG11668 Novel human diagno
15	160.5	14.5	185	24	ABW04008 Human expressed pr
16	160.5	14.5	185	24	ABW04009 Human expressed pr
17	160.5	14.5	185	24	ABW04010 Human expressed pr
18	160.5	14.5	185	24	ABW04011 Human expressed pr
19	160.5	14.5	185	24	ABW04012 Human expressed pr
20	160.5	14.5	185	24	ABW04014 Human expressed pr
21	160.5	14.5	185	24	ABW04015 Human expressed pr
22	149.5	13.5	419	24	ABW19109 pathogen specific
23	149.5	13.5	423	24	ABW18931 Homo sapiens secre
24	149	13.4	262	22	AAW95634 Homo sapiens secre
25	149	13.4	262	22	AAW24121 Human EST encoded
26	149	13.4	262	22	AAW38873 Human EST encoded
27	149	13.4	262	22	AAW27243 Human EST encoded
28	149	13.4	274	22	AAW40659 Human EST encoded
29	147.5	13.3	512	22	ABG00692 Novel human diagno
30	145	13.1	528	22	AAW81105 Mycobacterium tube
31	143	12.9	51	22	ABG27668 Novel human diagno
32	143	12.9	354	21	AAW44289 Arabidopsis thaila
33	141.5	12.7	368	21	AAW26449 Arabidopsis thaila
34	139	12.5	527	23	ABW05697 M. tuberculosis an
35	138	12.4	269	21	AAW95625 Arabidopsis thaila
36	135	12.2	369	21	AAW47270 Arabidopsis thaila
37	135	12.2	381	21	AAW47292 Arabidopsis thaila
38	135	12.2	594	22	ABW61362 Drosophila melanog
39	132.5	11.9	355	22	AAW40927 Human polypeptide
40	132	11.9	460	22	ABW63425 Drosophila melanog
41	131	11.8	381	21	AAW26698 Arabidopsis thaila
42	130.5	11.8	163	23	AAW14306 Sugarcane proline
43	127.5	11.5	707	22	AAW39141 Human polypeptide
44	126	11.4	1260	22	ABW59060 Drosophila melanog
45	125	11.3	204	20	AAW13498 Tissue cement prot

ALIGNMENTS

RESULT 1	AAW90961	AAW90961 standard; Protein; 195 AA.
XX	XX	AAW90961;
AC	XX	14-UTL-2000 (first entry)
XX	XX	Human CSGP-1 protein.
DE	XX	Human CSGP-1 protein.
XX	XX	Human; CSGP-1; cell surface glycoprotein; CSGP-2; anti-tumoric;
KW	XX	neurotropic; neuroprotective; anticonvulsant; gene therapy; treatment;
KW	XX	diagnosis; hematologic disorder; anemia; myeloproliferative disorder;
KW	XX	lymphoma; karyotypic disorder; Klinefelter syndrome; Turner syndrome;
KW	XX	neuronal disorder; akathisia; Alzheimer's disease; amnesia; catatonia;
KW	XX	epilepsy; neurofibromatosis.
OS	XX	Homo sapiens.
PN	XX	US6043056-A.
XX	XX	28-MAR-2000.
PD	XX	06-NOV-1996; 98US-0187331.
PF	XX	06-NOV-1998; 98US-0187331.
PR	XX	(INCY-) INCYTE PHARM INC.
PA	XX	Yue H, Gorgone GA, Baughn MR, Corley NC, Guegler KJ,
PI	XX	WPI, 2000-270341/23.
DR	XX	

XX DE Human EST encoded protein SEQ ID NO: 1434.
XX KM Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
XX KM tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
XX KM diagnostics; forensic test; gene mapping; genetic disorder;
XX KM biodiversity; gene therapy; nutrition.
XX OS Homo sapiens.
XX PN WO200154477-A2.
XX PD 02-AUG-2001.
XX PF 25-JAN-2001; 2001WO-US02687.
XX PR 25-JAN-2000; 2000US-0491404.
XX PR 17-JUL-2000; 2000US-0617746.
XX PR 03-AUG-2000; 2000US-0631451.
XX PR 15-SEP-2000; 2000US-0663870.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
XX PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX DR WPI; 2001-476164/51.
XX DR N-PSDB; AAS98568.
XX PT Isolated polypeptide for treatment of diseases, diagnostic, raising
XX PT antibodies and research use -
XX PS Claim 20; Page 999; 1275pp; English.
XX CC The present invention provides the protein and coding sequences of novel
XX CC proteins from a variety of organisms, including human, dog, cat, horse,
XX CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea urchin,
XX CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
XX CC from the organism of interest. They can be used in diagnostics,
XX CC forensics, gene mapping, identification of mutations, to assess
XX CC biodiversity and for nutritional purposes. The present sequence is a
XX CC protein of the invention.
XX SQ Sequence 168 AA;
SQ Query Match 79.8%; Score 886; DB 22; Length 168;
Best Local Similarity 89.4%; Pred. No. 3.5e-69;
Matches 161; Conservative 1; Mismatches 2; Indels 16; Gaps 2;
QY 1 MESWGLPCLAFICFLMHARGRDPDLADLDDPEPTKKNSDIYFKKPYTPQENPD 60
DB 1 MESWGLPCLAFICFLMHARGRDPDLADLDDPEPTKKNSDIYFKKPYTPQENPD 60
QY 61 SGGNIYRPRPPOPOGNSGNSGG-YFNDVDRDGRYPFRPRPAGGGGGSSYGN 119
DB 61 SGGNIYRPRPPOPOGNSGNSGGSYFNDVDRDGRYPFRPRPAGGGGGSSYGN 120
QY 120 SDNTHRGYRPNRSRYNTYGGDHSSTYGNPEGNMVAKIYPIVSVVVTLLGAASVFX 179
DB 121 SDNTH-----GGDHSSTYGNPEGNMVAKIYPIVSVVVTLLGAAGLQ 165
RESULT 4
ABG00690 ID ABG00690 standard; Protein; 96 AA.
XX AC ABG00690;
XX DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #681.
XX KM Human; chromosome mapping; gene mapping; gene therapy; forensic;

KM food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX DR WPI; 2001-639362/73.
XX DR N-PSDB; AAS64877.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -
XX PS Claim 20; SEQ ID No 31049; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantifying a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG0010-ABG0377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 96 AA;
SQ Query Match 45.2%; Score 502; DB 22; Length 96;
Best Local Similarity 96.8%; Pred. No. 3.7e-36;
Matches 92; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 101 PRPRPAGGGGGGSSYGNSDNTHRGYRPNRSRYNTYGGDHSSTYGNPEGNMVAKIY 160
DB 2 PRGPAPAGGGGGGSSYGNSDNTHRGYRPNRSRYNTYGGDHSSTYGNPEGNMVAKIY 61
QY 161 PIVSVVVTLLGAASVFXFLANRRNCFRTHERENV 195
DB 62 PIVSVVVTLLGAASVFXFLANRRNCFRTHERENV 96
RESULT 5
AAG00621 ID AAG00621 standard; Protein; 91 AA.
XX AC AAG00621;
XX DT 06-OCT-2000 (first entry)
XX DE Human secreted protein, SEQ ID NO: 4702.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.
 XX Homo sapiens.
 OS
 XX EP1033401-A2.
 PN
 PD 06-SEP-2000.
 XX
 PF 21-FEB-2000; 2000EP-0200610.
 XX
 XX 26-FEB-1999; 99US-0122487.
 FR
 XX (GSEST) GENSET.
 PA
 PI Dumas Malne Edwards J, Duclet A, Giordano J;
 XX WPI; 2000-500381/45.
 DR N-PSDB; AAC00627.
 XX
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 13; SEQ ID 4702; 71pp + CD-ROM; English.
 CC The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNA or polyA+ RNAs derived from 30
 CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (UTR) of the mRNA because they are often obtained
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 CC those cases where longer cDNA sequences have been obtained, the full 5'
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 CC ends and can therefore be used to obtain full length cDNAs and genomic
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.
 CC
 SQ Sequence 91 AA;
 Query Match 44.2%; Score 491; DB 21; Length 91;
 Best Local Similarity 98.8%; Pred. No. 3.2e-35;
 Matches 84; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MESWGLPCLAFCLPMHARGQDFDLADLDPEPTKKPNSDIYPKXPYYPPQENPD 60
 DB 1 MESWGLPCLAFCLPMHARGQDFDLADLDPEPTKKPNSDIYPKXPYYPPQENPD 60
 QY 61 SGNITYPKPRPQPPQPGNSGNSG 85
 DB 61 SGNITYPKPRPQPPQPGNSGNSG 85
 RESULT 6
 ABG27670
 ID ABG27670 standard; Protein; 345 AA.
 XX
 XX ABG27670;
 AC
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #27661.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 OS
 XX Homo sapiens.
 XX
 PN WO200175067-A2.
 XX

PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 FR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX WPI; 2001-639362/73.
 DR N-PSDB; AAS91857.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20; SEQ ID No 58029; 103pp; English.
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridization probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 345 AA;
 Query Match 41.9%; Score 465; DB 22; Length 345;
 Best Local Similarity 71.1%; Pred. No. 2.4e-32;
 Matches 86; Conservative 2; Mismatches 11; Indels 22; Gaps 2;
 QY 5 WGLPCLAFCLPMHARGQDFDLADLDPEPTKKPNSDIYPKXPYYPPQENPD 64
 DB 103 WRRPD-FTIRLTATGSGQDFDLADLDPEPTKKPNSDIYPKXPYYPPQENPD 161
 QY 65 IYPRPKPRPQPPQPGNSGSGYFNDVDRDGRYPRPRPPAGGGGGGYSYNSDPTH 124
 DB 162 -----YFNDVDRDGRYPRPRPPAGGGGGGYSYNSDPTH 200
 QY 125 G 125
 DB 201 G 201
 RESULT 7
 ABG27669
 ID ABG27669 standard; Protein; 91 AA.
 XX
 XX ABG27669;
 AC
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #27660.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW

KM food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 OS
 XX MO200175067-A2.
 PN
 XX 11-OCT-2001.
 PD
 XX 30-MAR-2001; 2001WO-US08631.
 PF
 XX 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX Drmanac RT, Liu C, Tang YT;
 PI
 XX WPI; 2001-639362/73.
 DR N-PSDB; AAS91856.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20; SEQ ID No 58028; 103bp; English.
 XX
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 91 AA;
 SQ

Query Match 38.5%; Score 427; DB 22; Length 91;
 Best Local Similarity 98.6%; Pred. No. 1.1e-29;
 Matches 72; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 35 EPIKKNSSIIYKPKPPYFPQPNPDSGGNIIYRPPRPQPOGNSGNGGFNDVDRD 94
 DB :|||||
 DB 19 KPIKKNSSIIYKPKPPYFPQPNPDSGGNIIYRPPRPQPOGNSGNGGFNDVDRD 78
 QY 95 GRYPFRPRPRPPA 107
 DB :|||||
 DB 79 GRYPFRPRPRPPA 91

RESULT 8
 ABG00689
 ID ABG00689 standard; Protein; 73 AA.
 AC ABG00689;
 XX
 XX 13-FEB-2002 (first entry)
 DT
 XX
 DE Novel human diagnostic protein #680.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 OS
 XX MO200175067-A2.
 PN
 XX 11-OCT-2001.
 PD
 XX 30-MAR-2001; 2001WO-US08631.
 PF
 XX 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX Drmanac RT, Liu C, Tang YT;
 PI
 XX WPI; 2001-639362/73.
 DR N-PSDB; AAS64876.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20; SEQ ID No 31048; 103bp; English.
 XX
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 73 AA;
 SQ

Query Match 33.5%; Score 371.5; DB 22; Length 73;
 Best Local Similarity 82.8%; Pred. No. 5.7e-25;
 Matches 72; Conservative 0; Mismatches 0; Indels 15; Gaps 1;

QY 109 GGGGGSSGYSNDYHGRGYPNSRYGNTYCGDHSHTYGNPEGMMVAKIVSPIVSVVV 168
 DB :|||||
 DB 2 GGGGGSSGYSNDYHGRGYPNSRYGNTYCGDHSHTYGNPEGMMVAKIVSPIVSVVV 46
 QY 169 TLIGAAASYFKLNNRNCERTHEPENV 195
 DB :|||||
 DB 47 TLIGAAASYFKLNNRNCERTHEPENV 73

RESULT 9
 ABG03178
 ID ABG03178 standard; Protein; 62 AA.
 AC ABG03178;
 XX
 XX 13-FEB-2002 (first entry)
 DT

XX Novel human diagnostic protein #3169.
XX Human chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX N-PSDB; AAS67365.
XX
XX Claim 20; SEQ ID No 33537; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 62 AA;
SQ
Query Match 27.7%; Score 308; DB 22; Length 62;
Best Local Similarity 98.1%; Pred. No. 1.5e-19;
Matches 53; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MESWAGPCLAFGLFMHARGQRFDLADLDEPTKPKNSDIYPRKPPYV 54
Db 1 MESWAGPCLAFGLFMHARGQRFDLADLDEPTKPKNSDIYPRKPPYV 54
RESULT 10
AAB19128
ID AAB19128 standard; Protein, 169 AA.
XX
XX AAB19128;
XX
XX 19-FEB-2001 (first entry)
XX
XX Polypeptide isolated from lymph node stromal cells of fan -/- mice.
DE

XX Lymph node stromal cell; fan -/- mice; inflammatory disorder;
XX immune system disorder; cancer; viral infection;
XX blood vessel growth; tumour necrosis factor disorder; arthritis;
XX inflammatory bowel disease; fibroblast growth factor-mediated disorder;
XX cardiac failure.
XX Mus sp.
XX OS
XX WO200058463-A1.
XX 05-OCT-2000.
XX 18-FEB-2000; 2000WO-NZ00015.
XX 25-MAR-1999; 99US-0276268.
XX 26-AUG-1999; 99US-0383586.
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX Strachan L, Sleeman M, Abernethy N, Onrust R, Kumble KD;
XX Murison JG;
XX WPI; 2000-664924/64.
XX N-PSDB; AAA96738.
XX
XX Polypeptide expressed in mammalian fan -/- lymph node stromal cells,
XX useful for modulating growth of blood cells, for treating inflammatory
XX and tumour necrosis factor-mediated disorders, cancer and viral
XX disorders -
XX
XX Claim 1; Page 71; 75pp; English.
XX
XX The present sequence represents a polypeptide sequence which is
XX isolated from lymph node stromal cells of fan -/- mice. The
XX polynucleotide and their polypeptides are useful for treating an
XX inflammatory disorder, disorder of immune system and cancer selected
XX from epithelial, lymphoid, myeloid, stromal and neuronal cancers, a
XX viral disorder, in particular HIV infection and for modulating the
XX growth of blood vessels. The polypeptides are useful for treating a
XX tumour necrosis factor (TNF) mediated disorder, such as those selected
XX from arthritis, inflammatory bowel disease and cardiac failure and a
XX fibroblast growth factor-mediated disorder. It is also useful in assays
XX to determine biological activity, to raise antibodies, to isolate
XX corresponding ligands or receptors, to quantify levels of protein or
XX cognate corresponding ligand or receptors, as anti-inflammatory agents,
XX and in compositions for the treatment of skin, connective tissue and
XX immune system diseases. The polynucleotide is useful as marker for
XX tissue, as a chromosome marker or tags in the identification of a
XX genetic disorder.
XX
XX Sequence 169 AA;
SQ
Query Match 15.9%; Score 176.5; DB 21; Length 169;
Best Local Similarity 29.6%; Pred. No. 1.1e-07;
Matches 59; Conservative 17; Mismatches 56; Indels 67; Gaps 10;
QY 10 LAFGLFMHARGQRFDLADLDEPTKPKNSDIYPRKPPYVPOPNPDSG----- 62
Db 11 LALLGAAGAASDDPDLADLDEP-----NMKPTKKA---PTPKPSGGFDLEAL 59
QY 63 -----GNTYPRKPRRPOPQNGNSGGYFNDVDXDDGYPFRPRRRPAG-G 109
Db 60 PGGGGGAGEKPGN---RPQDPKPR-PRPHGDSGG-ISDLDADA-----AGGG 103
QY 110 GGGGYSSYGSNDTHGRGGYRPNRSRYGNTYGGDHSTYGNPGNMAKIVSPVSVVVT 169
Db 104 GGAAGRGSGDEGGHGAAGAPB-----GTGPG-----LVPGVAIVAA 143
QY 170 LIGAAASYFKLNRRNCFR 188
Db 144 VAGAVSSFVAVOGRRRLCFR 162

QY 118 GNSDNTHGRGGRPNRSRYGNTYGGDHHSTYGNPEGNNVAKIVSPISVVVTLGLAAASY 177
 Db 121 GGSHRKEGEADAPG-----VTPGIVGAVVAAGALISF 155
 QY 178 FKLNRRNRCFR 188
 Db 156 IAYOKKKLCPK 166

RESULT 13
 ABU04013
 ID ABU04013 standard; Protein, 195 AA.
 AC ABU04013;
 DT 29-JAN-2003 (first entry)
 DE Human expressed protein tag (EPT) #679.
 KW Translational profiling; expressed protein tag; EPT; kinase;
 KW phosphatase; protease; protease inhibitor; transporter;
 KW cytoskeletal protein; receptor; transcription factor; cancer; MHC;
 KW major histocompatibility complex; myeloma; colon cancer;
 KW gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;
 KW leukaemia.
 XX Homo sapiens.
 OS
 PN WO200278524-A2.
 PD 10-OCT-2002.
 DE 28-MAR-2002; 2002WO-US09671.
 PF 28-MAR-2001; 2001US-279495P.
 PR 21-MAY-2001; 2001US-292544P.
 PR 08-AUG-2001; 2001US-310801P.
 PR 01-OCT-2001; 2001US-326370P.
 PR 04-DEC-2001; 2001US-336780P.
 PR 20-FEB-2002; 2002US-358985P.
 XX (ZYCO-) ZYCO INC.
 PA
 XX Chicx RM, Tomlinson AJ, Urban RG;
 XX WPI; 2003-040607/03.
 DE New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
 PT cytoskeletal proteins, receptors or transcription factors), useful for
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma
 PT or leukemia -
 XX Example 2; SEQ ID No 679; 134pp; English.

The invention describes a purified polypeptide, which comprises a
 fragment of a kinase, phosphatase, protease, protease inhibitor,
 transporter, cytoskeletal protein, receptor or transcription factor.
 CC The polypeptide is useful as an immunogenic composition for eliciting
 CC in a mammal an immunogenic response directed against any of the purified
 CC polypeptide. The purified polypeptide, or the antibody that binds to
 CC this polypeptide, is useful for treating cancer. The polypeptide is
 CC also useful for identifying compounds that binds to a naturally
 CC processed class I or class II MHC-binding polypeptide. The polypeptides
 CC and polynucleotides are particularly useful for treating or preventing
 CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
 CC lymphoma or leukemia. These are also useful for screening agents for
 CC treating the above mentioned diseases. This sequence represents an
 CC expressed protein tag (EPT) isolated from human tissue for translational
 CC profiling.
 CC Note: This sequence does not appear in the printed specification but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

SEQ Sequence 195 AA;
 Query Match 14.8%; Score 164.5; DB 24; Length 195;
 Best Local Similarity 28.3%; Pred. No. 1.5e-06;
 Matches 54; Conservative 14; Mismatches 56; Indels 67; Gaps 7;

QY 15 FLHARAGQRDLDALDDEPTKKPNSDIYPRKPPY-----PQPE 57
 Db 26 FLHGERNNGGFDLSDALPDNE-NKKPTA-----IPKKEAGDDPDLGDAVVDGENDDPRPP 80
 QY 58 NPDSGNIYPRPRPRQPOGNGSGNGGYNDVDRDGRPRPRPRPPAGGGGGYSS 117
 Db 81 NP-----PKPMNPNP-NHPSSSGSPSDADLDG-----VSGSGKGGSDG 120
 QY 118 GNSDNTHGRGGRPNRSRYGNTYGGDHHSTYGNPEGNNVAKIVSPISVVVTLGLAAASY 177
 Db 121 GGSHRKEGEADAPG-----VTPGIVGAVVAAGALISF 155
 QY 178 FKLNRRNRCFR 188
 Db 156 IAYOKKKLCPK 166

RESULT 14
 ABG11668
 ID ABG11668 standard; Protein, 265 AA.
 AC ABG11668;
 DT 18-FEB-2002 (first entry)
 DE Novel human diagnostic protein #11659.
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 OS Homo sapiens.
 PN WO200175067-A2.
 PD 11-OCT-2001.
 DE 30-MAR-2001; 2001WO-US08631.
 PF 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSR-) HYSR INC.
 PA
 XX Dimaac RT, Liu C, Tang YT;
 XX N-PSDB; AAS75855.
 DR WPI; 2001-639362/73.
 DE New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX Claim 20; SEQ ID No 42027; 103pp; English.

The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG039377 represent novel human
CC amino acid sequences. The sequence data for this patent did not appear in the printed
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.
CC

SQ Sequence 265 AA;

Query Match	14.6%	Score 162.5;	DB 22;	Length 265;
Best Local Similarity	25.5%;	Pred. No. 3.1e-06;		
Matches 59;	Conservative 15;	Mismatches 60;	Indels 97;	Gaps 91;

QY WGLRCLALFCLFMH-----ARG-----QRDTLADALDDP 34
Db WGIICPAFAHAIHSGTVPARGSGRTWARGAALLALLFGLLGVLYAAPDGGDLSLDALPDN 94
QY EPRKKDNDSDIYPRKRPYY-----POPNPDGGGINTYPRKRPPOP 77
Db E-NKKF7A-----IPKKSAGDDPDLGDAVVDGENDPPRPNP-----PKMPPNP 140
QY GAGGNGGGYINDVDRDGRYPAPRPRPPRPGCGGGCGYGSYNGSDNTHGRGGYRNSRYGN 137
Db 141 -NHSSSGGSFDLADLGD-----VSGEGRGKGDGCGSHRKEGEAEADAG----- 184
QY 138 TYGGDHHSTYGNDEGNVAKIVSPITSVYVYVLLLAASLYFLNNRRRCFR 188
Db 185 -----VIRGIVGAVVAAVQAISFLAYQKKKLCFK 215

RESULT 15

ID ABU04008 standard; Protein; 185 AA.

AC ABU04008;

DT 29-JAN-2003 (first entry)

Human expressed protein tag (EPT) #674.

KM Translational profiling; expressed protein; Epi; Kinase;
KM Phosphatase; protease; protease inhibitor; transporter;
KM cytoskeletal protein; receptor; transcription factor; cancer; MHC;
KM major histocompatibility complex; myeloma; colon cancer;
KM gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;
KM leukaemia.

Homo sapiens.

PN W0200278524-A2.

PD 10-OCT-2002.

PF 28-MAR-2002; 2002WO-US09671.

PR 28-MAR-2001; 2001US-279495P
PR 21-MAY-2001; 2001US-292544P
PR 08-AUG-2001; 2001US-310801P
PR 01-OCT-2001; 2001US-326370P
PR 04-DEC-2001; 2001US-336780P
PR 20-FEB-2002; 2002US-358985P

PA (ZYCO-) ZYCOS INC.

PI Chicz RM, Tomlinson AJ, Urban RG;

DR WPI; 2003-040607/03.

PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters, cytoskeletal proteins, receptors or transcription factors), useful for

PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma
 PT or leukemia -
 XX
 PS Example 2; SEQ ID No 674; 134pp: English.
 PV

The invention describes a purified polypeptide, which comprises a fragment of a kinase, phosphatase, protease, protease inhibitor, transporter, cytoskeletal protein, receptor or transcription factor. The polypeptide is useful as an immunogenic composition for eliciting in a mammal an immunogenic response directed against any of the purified polypeptide. The purified polypeptide, or the antibody that binds to this polypeptide, is useful for treating cancer. The polypeptide is also useful for identifying compounds that binds to a naturally processed class I or class II MHC-binding polypeptide. The polypeptides and polynucleotides are particularly useful for treating or preventing myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma lymphoma or leukemia. These are also useful for screening agents for treating the above mentioned diseases. This sequence represents an expressed protein tag (EPT) isolated from human tissue for translational profiling.

Note: This sequence does not appear in the printed specification but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 185 AA;

Query Match	14.5%;	Score 160.5;	DB 24;	Length 185;
Best Local Similarity	-28.0%;	Pred. No. 3.1e-06;		
Matches 56;	Conservative 16;	Mismatches 57;	Indels 71;	Gaps 8;

```

QY 10 LALPLC-----LNMHAGQGRFDLADLDLPBEPTKKPNSDIYKPKPPYY-----53
Dd 7 LALLLEFGLLGLVLAAPDDGFFDLSDLPLPNE--NKETPA-----IPKKEPSAGDDFDLGDAAVD 61
QY 54 -----EQENPDPSGNITYPRPKRPQOPQPSGNSGSGYFNDVDKDDRGYPRPAPRPAPG 108
Dd 62 GENDDDRPNP-----PKPMWNPMP--NHPSSSGSFSDALDAG-----VSG 101
QY 109 GGGGGYSSISGNSDNTHGKGYRPNRSRYNTYGGDHSITGNPEGNMVAKIVSPVSVVVV 168
Dd 102 GEEKSGSDGGSGHRRKEGEADAPG-----VIPGIVGAVVV 136
QY 169 TELGAAASYFKLNNRNCFR 188
Dd 137 AVAGATSSFTIAYQKKCLCFK 156

```

Search completed: February 18, 2004, 19:05:13
Job time : 42 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2004, 19:06:36 ; Search time 34 Seconds

(without alignments)
1200.869 Million cell updates/sec

Title: US-09-997-701-1

Perfect score: 1110
1 MESMWGLPCLAFICFLMAR.....SYFKLNRRNCRTHEPENV 195

Sequence: BLOSUM62

Scoring table: GAP0 10.0 , Gapext 0.5

Searched: 801455 seqs, 209382283 residues

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/prodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/prodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/prodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/prodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/prodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1110	100.0	195	10	US-09-997-701-1
2	1004.5	90.5	180	10	US-09-997-701-5
3	176.5	15.9	169	9	US-09-823-038A-53
4	172.5	15.5	165	11	US-09-866-050A-638
5	164.5	14.8	155	9	US-09-925-302-717
6	160.5	14.5	185	12	US-10-331-496A-24
7	149	13.4	262	11	US-09-746-783-65
8	139	12.5	527	10	US-09-712-363-156
9	139	12.5	527	16	US-10-080-170-348
10	138	12.4	229	11	US-09-746-783-38
11	127.5	11.4	707	11	US-09-919-093-278
12	127	11.4	1046	15	US-10-156-761-10088
13	125	11.3	203	15	US-10-226-489-6
14	125	11.3	204	12	US-10-280-114-9
15	121.5	10.9	488	16	US-10-080-170-4

16	121.5	10.9	633	14	US-10-086-464-11	Sequence 11, Appl
17	121	10.9	455	12	US-10-032-585-7399	Sequence 7399, Ap
18	119	10.7	1004	10	US-09-738-626-5676	Sequence 5676, Ap
19	118.5	10.7	264	12	US-10-304-630-24	Sequence 24, Appl
20	117	10.5	922	10	US-09-855-754-6	Sequence 6, Appl
21	117	10.5	922	12	US-10-302-896-6	Sequence 6, Appl
22	117	10.5	922	15	US-10-227-353-6	Sequence 6, Appl
23	117	10.5	953	9	US-09-888-615-66	Sequence 66, Appl
24	116.5	10.5	940	12	US-10-291-172-363	Sequence 363, App
25	116.5	10.5	968	12	US-10-291-172-739	Sequence 739, App
26	115	10.4	264	12	US-10-301-488A-30	Sequence 30, Appl
27	115	10.4	264	12	US-10-301-488A-33	Sequence 33, Appl
28	115	10.4	264	12	US-10-410-907A-13	Sequence 13, Appl
29	115	10.4	264	12	US-10-346-190-80	Sequence 80, Appl
30	115	10.4	264	12	US-10-417-964A-19	Sequence 19, Appl
31	115	10.4	264	12	US-10-417-964A-32	Sequence 32, Appl
32	115	10.4	264	15	US-10-209-194-2	Sequence 2, Appl
33	114.5	10.3	643	10	US-09-801-368-336	Sequence 236, App
34	113	10.2	251	12	US-10-029-386-33860	Sequence 33860, A
35	113	10.2	312	15	US-10-156-761-10983	Sequence 10983, A
36	110.5	10.0	256	12	US-10-304-630-31	Sequence 31, Appl
37	110.5	10.0	258	12	US-10-301-488A-27	Sequence 27, Appl
38	110.5	10.0	264	12	US-10-304-630-27	Sequence 27, Appl
39	110	9.9	54	10	US-09-855-754-22	Sequence 22, Appl
40	110	9.9	54	12	US-10-302-896-22	Sequence 22, Appl
41	109	9.8	1228	11	US-09-917-384-1	Sequence 1, Appl
42	109	9.8	1228	11	US-09-917-383-1	Sequence 1, Appl
43	108.5	9.8	256	12	US-10-304-630-25	Sequence 25, Appl
44	108.5	9.8	256	12	US-10-304-630-26	Sequence 26, Appl
45	108.5	9.8	256	12	US-10-304-630-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1	US-09-997-701-1
Sequence 1, Application US/09997701	
Patent No. US20020107180A1	
GENERAL INFORMATION:	
APPLICANT: Yue, Henry	
APPLICANT: Corley, Neil C.	
APPLICANT: Guejter, Karl J.	
APPLICANT: Gorgone, Gina A.	
APPLICANT: Baughn, Mariah R.	
TITLE OF INVENTION: CELL SURFACE GLYCOPROTEINS	
FILE REFERENCE: PF-0631 US	
CURRENT APPLICATION NUMBER: US/09/997,701	
CURRENT FILING DATE: 2001-11-30	
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/470,946	
PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-22	
NUMBER OF SEQ ID NOS: 6	
SOFTWARE: PERL Program	
SEQ ID NO 1	
LENGTH: 195	
TYPE: PRT	
ORGANISM: Homo sapiens	
FEATURE:	
OTHER INFORMATION: 2297891	
US-09-997-701-1	
Query Match	100.0%; Score 1110; DB 10; Length 195;
Best Local Similarity	100.0%; Pred. No. 2.1e-87;
Matches 195; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MESMWGLPCLAFICFLMARAGGRDPLADALDDPEPTKKKPNSDIYKPKPPYQPEPND 60
DB	1 MESMWGLPCLAFICFLMARAGGRDPLADALDDPEPTKKKPNSDIYKPKPPYQPEPND 60
QY	61 SGNITYPKPKPPQPPGNSGSGYFNDVDRDGRYPRPRPRPAGGGGGGYSYGS 120
DB	61 SGNITYPKPKPPQPPGNSGSGYFNDVDRDGRYPRPRPRPAGGGGGGYSYGS 120

Qy	121	DNTHGRGQYAPNRYNTGTGGDHSHTYGNPEGNMVAKTVPISVSVVVTTLGGAASYFKL	180
	121	DNTHGRGQYAPNRYNTGTGGDHSHTYGNPEGNMVAKTVPISVSVVVTTLGGAASYFKL	180
Db	121	DNTHGRGQYAPNRYNTGTGGDHSHTYGNPEGNMVAKTVPISVSVVVTTLGGAASYFKL	180
Qy	181	NNRRNCRTHEPENV	195
	181	NNRRNCRTHEPENV	195
Db	181	NNRRNCRTHEPENV	195

RESULT 2
US-09-997-701-5

Query Match %	90.5%	Score 1004.5	DB 10	Length 180
Best Local Similarity	92.3%	Pred. No. 2.1e-78		
Matches 180, Conservative	0	Mismatches 0	Indels 15	Gaps 1

Qy	61	SGNGIYRPRPRQPPQPGNSGNSGGYFNVVDVDDGAPRPRPRPPAAGGGGGYISVGN	120
Db	61	SGNGIYRPRPRQPPQPGNSGNSGGYFNVVDVDDGAPRPRPRPRPPAAGGGGGYISVGN	120
Qy	121	DNTHGRCGYPRNRYGNTYGGDHSHTSGNPEGNNIAKIVSPVSYVYVYTTTIGAAASFKL	180
Db	121	DNTH-----GGDHSHTSGNPEGNNIAKIVSPVSYVYVYVYTTTIGAAASFKL	165

```

, RESULT 3
, US-09-823-038A-53
, Sequence 53, Application US/09823038A
, Patent No. US20020058335A1
, GENERAL INFORMATION:
, APPLICANT: Strachan, Lorna
, APPLICANT: Sleeman, Matthew
, APPLICANT: Abernethy, Nevil
, APPLICANT: Ormrod, Rene
, APPLICANT: Kumble, Anand
, TITLE OF INVENTION: Compositions Isolated From Stromal Cells
, TITLE OF INVENTION: and Methods For Their Use
, FILE REFERENCE: 11000.1037c3
, CURRENT APPLICATION NUMBER: US/09/823,038A
, CURRENT FILING DATE: 2001-07-09
, NUMBER OF SEQ ID NOS: 61

```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 53
;
; LENGTH: 169
;
; TYPE: PRF
;
; ORGANISM: Mouse
US-09-823-038A-53

```

Query Match 15.9%; Score 176.5; DB 9; Length 169;
Best Local Similarity -29.6%; Pred. No. 1.4e-07;
Matches 59; Conservative 17; Mismatches 56; Indels 67; Gaps 10;

Qy	63	-----GNIYPRKRPQPQPNSSGSGFYFDVDRDDGRYPRPRPPRG-G	109
Db	60	PGGGGGAAGEKPGN---RPQDPKR-PRPHDSGG-ISDSDLAIA-----AGGC	103
Qy	110	GGGGYSYNSNDTHGRGGRGVRPNSSKYGNTYGGDHHSTYGNBGMVAKIVSPISVVVVT	169
Db	104	CGAGRGSGDSBEGHGAGAGEP-----GTQG-----LVPGVLAAYVA	143
Qy	170	LLGAASFYKLNRRRCR	188
Db	144	VAGAVSFWAYQRRRCR	162

```

1  RESULT 4
2  US-09-866-050A-638
3  : Sequence 638: Application US/09866050A
4  : Publication No. US20030040471A1
5  :
6  : GENERAL INFORMATION:
7  :
8  : APPLICANT: Watson, James D.
9  : APPLICANT: Strachan, Lorna
10 : APPLICANT: Sleeman, Matthew
11 : APPLICANT: Onrust, Rene
12 : APPLICANT: Murison, James G.
13 : APPLICANT: Kumble, Krishnaand D.
14 :
15 : TITLE OF INVENTION: Compositions for Their
16 : TITLE OF INVENTION: and Methods for Their
17 :
18 : FILE REFERENCE: 11000.1011c4U
19 :
20 : CURRENT APPLICATION NUMBER: US/09/866.050A
21 :
22 : CURRENT FILING DATE: 2001-05-24
23 :
24 : NUMBER OF SEQ ID NOS: 725
25 :
26 : SOFTWARE: FastSeq for Windows Version 4.0
27 :
28 : SEQ ID NO 638
29 :
30 : LENGTH: 165
31 :
32 : TYPE: PR1
33 :
34 : ORGANISM: Rat
35 :
36 : US-09-866-050A-638

```

Query Match	15.5%	Score 172.5;	DB 11;	Length 165;
Best Local Similarity	31.2%	Pred. No. 2.9e-07;		
Matches 60; Conservative	18;	Mismatches 55;	Indels 59;	Gaps 9;

QY 62 GGNIVP-----RPKRRPQPGNGNSGGYNDVDNRDGRPPRRPRRRPAGGGGGYSS 116
DB 65 GAAEDRRRPGSPKPDPKP-PGPPDSSGG-1SDDLBLE-----VAGHGGR-- 108
QY 117 YGNSDWTGRRGGYRPNRSRGNTYGGDHSSTYGNGPEGNMVAKIYSPISVVVVTLLGAAS 176
DB 109 -----GGAADR-----GTDGAESGGPQG-----LIPGVAAVTLAALAGVSS 146
QY 177 YFKLANRRRCNFR 188
DB 147 FVAYGRRLCFR 158

RESULT 5
US-09-925-302-717
Sequence 717, Application US/09925302
Patent No. US20020044941A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA104
CURRENT APPLICATION NUMBER: US/09/925,302
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05918
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 896
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 717
LENGTH: 195
TYPE: PRT
ORGANISM: Homo sapiens
US-09-925-302-717

Query Match 14.8%; Score 164.5; DB 9; Length 195;
Best Local Similarity 28.3%; Pred. No. 1,7e-06;
Matches 54; Conservative 14; Mismatches 56; Indels 67; Gaps 7;
QY 15 FLHARGQDFDLADLDPEPTKPNDSIYPRKPPY-----POPE 57
DB 26 FLGGERNNGGFDLSDALPDNE-NKKPTA-----IPKPSAGDDPDLGDVAVDGENDP RRP 80
QY 58 NPDSGNIYPRKPPPOPOGNSGSGYFNDVDRDGRPRPRPRPPGCGGCGYSS 117
DB 81 NP-----PKMPNP-NHPSSGSGSDADLDG-----VSGGSGKGSDDG 120
QY 118 GNSDNTHGSGYRPNRNTYGGDHSHTYGNPEGNVAKIVSPISVVVTLGAASY 177
DB 121 GSHRKEGEADAPG-----VIRGIVGAVVAVVAGAISSP 155
QY 178 FKLNRRNCR 188
DB 156 IAYQKKLCPK 166

RESULT 6
US-10-331-496A-24
Sequence 24, Application US/10331496A
Publication No. US20030228305A1
GENERAL INFORMATION:
APPLICANT: FRANTZ, GRETCHEN
APPLICANT: HILLAN, KENNETH J.
APPLICANT: PHILLIPS, HEIDI S.
APPLICANT: POLAKIS, PAUL
APPLICANT: SMITH, VICTORIA
APPLICANT: SPENCER, SUSAN D.
APPLICANT: WILLIAMS, P. MICKEY
APPLICANT: WU, THOMAS D.
APPLICANT: ZHANG, ZHEN
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
FILE REFERENCE: P5014R1-PCT
CURRENT APPLICATION NUMBER: US/10/331,496A
CURRENT FILING DATE: 2002-12-30
PRIOR APPLICATION NUMBER: US 60/345,444
PRIOR FILING DATE: 2002-01-02
PRIOR APPLICATION NUMBER: US 60/351,885
PRIOR FILING DATE: 2002-01-25
PRIOR APPLICATION NUMBER: US 60/360,066
PRIOR FILING DATE: 2002-02-25
PRIOR APPLICATION NUMBER: US 60/362,004
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/366,869
PRIOR FILING DATE: 2002-03-20
PRIOR APPLICATION NUMBER: US 60/366,284

PRIOR FILING DATE: 2002-03-21
PRIOR APPLICATION NUMBER: US 60/368,679
PRIOR FILING DATE: 2002-03-28
PRIOR APPLICATION NUMBER: US 60/404,809
PRIOR FILING DATE: 2002-08-19
PRIOR APPLICATION NUMBER: US 60/405,645
PRIOR FILING DATE: 2002-08-21
NUMBER OF SEQ ID NOS: 95
SEQ ID NO 24
LENGTH: 185
TYPE: PRT
ORGANISM: Homo sapien
US-10-331-496A-24

Query Match 14.5%; Score 160.5; DB 12; Length 185;
Best Local Similarity 28.0%; Pred. No. 3.5e-06;
Matches 56; Conservative 16; Mismatches 57; Indels 71; Gaps 8;
QY 10 LAFLCF-----LHARGQDFDLADLDPEPTKPNDSIYPRKPPY----- 53
DB 7 LALLFLGLGVVAAPDGGFDLSDALPDNE-NKKPTA-----IPKPSAGDDPDLGDVAVD 61
QY 54 -----POPENPDSGNIYPRKPPPOPOGNSGSGYFNDVDRDGRYPRPRPPAG 108
DB 62 GENDDP RPNP-----PKMPNP-NHPSSGSGSDADLDG-----VSG 101
QY 109 GGGGYSYSGNSDNTHGSGYRPNRNTYGGDHSHTYGNPEGNVAKIVSPISVVV 168
DB 102 GSGKGSDDGGSHRKEGEADAPG-----VIRGIVGAVV 136
QY 169 TLLGAASYFKLNRRNCR 188
DB 137 AVAGAISSPFIAYQKKLCPK 156

RESULT 7
US-09-746-783-65
Sequence 65, Application US/09746783
Publication No. US20030044935A1
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
McCoy, John M.
Lavallie, Edward R.
Racie, Lisa A.
Treacy, Maurice
Spaulding, Vikki
Agostino, Michael J.
Howes, Steven H.
Fechtel, Kim
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
ENCODING THEM
NUMBER OF SEQUENCES: 231
CORRESPONDENCE ADDRESS:
ADDRESSER: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/746,783
FILING DATE: 21-Dec-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mlaabincic, Debra J.
REGISTRATION NUMBER: 46,931
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400

TELEFAX: (617) 742-4214
 INFORMATION FOR SEQ ID NO: 65:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 262 amino acids
 TYPE: amino acid
 STRANDEDNESS: <unknown>
 MOLECULE TYPE: protein
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 65:
 US-09-746-783-65

Query Match 13.4%; Score 149; DB 11; Length 262;
 Best Local Similarity 25.4%; Pred. No. 4.9e-05;
 Matches 60; Conservative 29; Mismatches 91; Indels 56; Gaps 11;

QY 1 MESWNG--LPCIAFLCFIMHARGDRPD--LADALDPEPTKKP--NSDIYPRKPPY 53
 DB 1 MWAMSAFLVCLASLATLVGRSGDFDNLADYKSTSVKQMDHTTTTNRGETT 60
 QY 54 POPENPDGSG-----NIVPRKPPPOPOPGNSGSGYFN 88
 DB 61 RAPAPPGSGLDLADLDDDDGRKPGIGRERNHVTTKQPVTRAPANTLANDFD 120
 QY 89 --DV-----DRDDGYPRPPRPPAGGGGSGYSGSDNTHRGGRYRNSRYGN-TYG 140
 DB 121 LADALDNDNDHDDGR-----RKEIAGGG--FSDKDLIEDIVGGGEYKPKGKGDGRYG 171
 QY 141 GDHSTYGNPEGNVAK--IVSPYVSVVVTLLGAAASYFLNRRNCFTHEPEN 194
 DB 172 SN-----DDPGSMVAEPGTIAGVASALMALIGVSSYISYQKKCFPSIQGLN 222

RESULT 8
 US-09-712-363-156
 Sequence 156, Application US/09712363
 Patent No. US20020164588A1
 GENERAL INFORMATION:

APPLICANT: Bisenberg, David
 APPLICANT: Rotstein, Sergio H.
 APPLICANT: Marcotte, Edward M.
 TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
 FILE REFERENCE: 07419-032001
 CURRENT APPLICATION NUMBER: US/09/712.363
 PRIOR FILING DATE: 2000-11-13
 PRIOR APPLICATION NUMBER: PCT/US00/02246
 PRIOR FILING DATE: 2000-01-28
 PRIOR APPLICATION NUMBER: 60/179,531
 PRIOR FILING DATE: 2000-02-01
 PRIOR APPLICATION NUMBER: 60/117,844
 PRIOR FILING DATE: 1999-01-29
 PRIOR APPLICATION NUMBER: 60/118,206,
 PRIOR FILING DATE: 1999-02-01
 PRIOR APPLICATION NUMBER: 60/126,593
 PRIOR FILING DATE: 1999-03-26
 PRIOR APPLICATION NUMBER: 60/134,093
 PRIOR FILING DATE: 1999-05-14
 PRIOR APPLICATION NUMBER: 60/134,092
 PRIOR FILING DATE: 1999-05-14
 PRIOR APPLICATION NUMBER: 60/165,124
 PRIOR FILING DATE: 1999-11-12
 PRIOR APPLICATION NUMBER: 60/165,086
 PRIOR FILING DATE: 1999-11-12
 NUMBER OF SEQ ID NOS: 292
 SOFTWARE: FaastSeq for Windows Version 4.0
 SEQ ID NO 156
 LENGTH: 527
 TYPE: PRT
 ORGANISM: Mycobacterium tuberculosis
 US-09-712-363-156

Query Match 12.5%; Score 139; DB 10; Length 527;
 Best Local Similarity 29.4%; Pred. No. 0.00075;

Matches 48; Conservative 15; Mismatches 68; Indels 32; Gaps 9;

QY 18 HARGDRFDLADLDDPEPTKKNSDIYK----PKPPYPOPEPDGSGNIVPR----- 68
 DB 179 YARPQED---PRGGPPDGGSDPRGGYPRPTGTGYPQGRPRHNDQGD--YPEQIGYR 233
 QY 69 -----PKRPPQPGNSGSGYFND---VDRDDGRYPRPPRPPAG--GGGGYSSYG 118
 DB 234 DGGGYPRGRYPRGRYPRDGRYDQGRYPDGGGGYPPYPRQRPPVSPGPAAGYCAPG 293
 QY 119 NSDNTHGRGGRYRNSRYGNT-YGDDHSTYG-----NPEGNV 155
 DB 294 YDQGYRSGGYGSPSGGQPGYGG--YGEYGRGPARHREGSYV 334

RESULT 9
 US-10-080-170-348
 Sequence 348, Application US/10080170
 Publication No. US20030129601A1
 GENERAL INFORMATION:
 APPLICANT: COLE, S.T.
 TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
 IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
 TREATMENT OF MYCOBACTERIOSES
 FILE REFERENCE: 03495.0218
 CURRENT APPLICATION NUMBER: US/10/080,170
 PRIOR FILING DATE: 2002-06-10
 PRIOR APPLICATION NUMBER: 60/270,123
 PRIOR FILING DATE: 2001-02-22
 NUMBER OF SEQ ID NOS: 652
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 348
 LENGTH: 527
 TYPE: PRT
 ORGANISM: Mycobacterium tuberculosis
 US-10-080-170-348

Query Match 12.5%; Score 139; DB 16; Length 527;
 Best Local Similarity 29.4%; Pred. No. 0.00075;
 Matches 48; Conservative 15; Mismatches 68; Indels 32; Gaps 9;

QY 18 HARGDRFDLADLDDPEPTKKNSDIYK----PKPPYPOPEPDGSGNIVPR----- 68
 DB 179 YARPQED---PRGGPPDGGSDPRGGYPRPTGTGYPQGRPRHNDQGD--YPEQIGYR 233
 QY 69 -----PKRPPQPGNSGSGYFND---VDRDDGRYPRPPRPPAG--GGGGYSSYG 118
 DB 234 DGGGYPRGRYPRGRYPRDGRYDQGRYPDGGGGYPPYPRQRPPVSPGPAAGYCAPG 293
 QY 119 NSDNTHGRGGRYRNSRYGNT-YGDDHSTYG-----NPEGNV 155
 DB 294 YDQGYRSGGYGSPSGGQPGYGG--YGEYGRGPARHREGSYV 334

RESULT 10
 US-09-746-783-38
 Sequence 38, Application US/09746783
 Publication No. US20030044935A1
 GENERAL INFORMATION:
 APPLICANT: Jacobs, Kenneth
 McCoy, John M.
 Lavallie, Edward R.
 Racie, Lisa A.
 Treacy, Maurice
 Spaulding, Vikki
 Agoestino, Michael J.
 Howe, Steven H.
 Rechtel, Kim
 TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
 ENCODING THEM
 NUMBER OF SEQUENCES: 231
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genetics Institute, Inc.

```

STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/746,783
FILING DATE: 21-Dec-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Milasincic, Debra J.
REGISTRATION NUMBER: 46,931
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 229 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-09-746-783-38

Query Match      12.4%; Score 138; DB 11; Length 229;
Best Local Similarity 27.3%; Pred. No. 0.00037;
Matches 50; Conservative 24; Mismatches 63; Indels 46; Gaps 9;

QY 26 DLADLDDPEPTKKNSDIYKPKRPYYPQENPDGSGN---NLYPRKRPQPPQNGSG 81
Db 39 DLADLDDQDDGRK-----KFGIGRRRMNVTITTKRPVTPTRAPAN 80
QY 82 NSGGYFN-DV-----DRDGRYPYRPPRPPAGGGGGYSSYGNSDNTHGRGGYRPNR 134
Db 81 TLGNDFLDALDLDNRDNRDGR-----RKPIAGGGG-FSDKDLSDIYGSGEYKPDKG 131
QY 135 YGN-TYGGDHSHTGNPEGNNVAK--IVSPIVSVVVTLLGAASPTKLANRNCRTTHE 191
Db 132 KGGRRYGSN-----DDGSGVVAEPGTIAGVASALMALIGAVSSYSYQKKCFESIQ 186
QY 192 PEN 194
Db 187 GLN 189

RESULT 11
US-09-919-039-278
; Sequence 278, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; PRIOR FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 278
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030108871A1 2457215CD1
US-09-919-039-278

```

```

Query Match      11.5%; Score 127.5; DB 11; Length 707;
Best Local Similarity 28.6%; Pred. No. 0.0099;
Matches 42; Conservative 5; Mismatches 57; Indels 43; Gaps 6;

QY 34 PEPTKKNSDI-----YKPKRPYYPQENPDGSGN---NLYPRKRPQPPQNGSG 81
Db 147 PGTPTPPPPAVTSAPPGADPPPTPSSGVTTPPQAGPPPPPAVPGPGPKGPGPG 206
QY 82 NSGGYNDVDRDGRYPYRPPRPPAGGGGGYSSYGNSDNTHGRGGYRPNRNYGNTYGG 141
Db 207 PKG-----GKMPGPKP-----GGPGUSTGHPKPPHKGGBERG-----GR 245
QY 142 DHRSTY-----GNPEGNNVAKI 158
Db 246 QHHPYHQHGGPPPGPGGRSEKI 272

RESULT 12
US-10-156-761-10088
; Sequence 10088, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10088
; LENGTH: 1046
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10088

Query Match      11.4%; Score 127; DB 15; Length 1046;
Best Local Similarity 34.1%; Pred. No. 0.017;
Matches 45; Conservative 8; Mismatches 65; Indels 14; Gaps 7;

QY 28 ADALDDPEPTKKNSDIYKPKR--PPYYPQENPDGSGNLYPRKRPQ--PPQNGSGNG 84
Db 244 AGAPGGPRPQAGAGGDBARPRPQGGPGAGAPPTGGMPPQAPRPG--GAPG 301
QY 85 GYFNDVDR-DGGRYPYRPPR--RPPAGGGGGYSSYGNSDNTHGRGGYR--NSRYGNTY 140
Db 302 G-----NRPNPQMPQRPAPGPRPGGGGPGGGRPGGGRPGGGRPGGGRPGGGRPG 356
QY 141 GDHSTYGNPEG 152
Db 357 GGGGGPAGRPGG 368

RESULT 13
US-10-226-489-6
; Sequence 6, Application US/10226489
; Publication No. US20030078390A1
; GENERAL INFORMATION:
; APPLICANT: Guido Christiaan Paesen, Patricia Anne Nuttall
; TITLE OF INVENTION: Tissue Cement
; FILE REFERENCE: 2488-1-001
; CURRENT APPLICATION NUMBER: US/10/226,489
; PRIOR FILING DATE: 2002-08-23
; PRIOR APPLICATION NUMBER: US/09/554,547
US-10-226-489-6

```

;; PRIOR FILING DATE: 2000-07-07
;; NUMBER OF SEQ ID NOS: 17
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 6
;; LENGTH: 203
;; TYPE: PRT
;; ORGANISM: R. appendiculatus
US-10-226-489-6

Query Match 11.3%; Score 125; DB 15; Length 203;
Best Local Similarity 29.1%; Pred. No. 0.0043;
Matches 50; Conservative 10; Mismatches 64; Indels 48; Gaps 7;

QY 7 LPCAFLCLMHARGQDFDLADLDDEPTKKPNSDIYKPK-----KP 50
DB 11 LAFAFLLSAIIHNNQSCVDAA-----PTRRP-----MSPPGCAGPCFTGIATLLRP 59
QY 51 PYYQPEPNDGSGNITPRPKRPPQPGNSGSGYFNDVDRDGRYP-----PRP 101
DB 60 GGGQOPGGQOPGQGRP-PMRPPGPVPGTSGSPQGRPGAPR-PGPVGTSGSPQGRPNA 117
QY 102 RPRPPAGGGGGSSYGNSDNTHGGRGYPNSRYGNTYGGDHHSTYGNPEGN 153
DB 118 RPRPGPVGTPVSSPGSS-----PGSSPGISIGTPLGTPLGTPQGS 159

RESULT 14
US-10-280-114-9
;; Sequence 9, Application US/10280114
;; Publication No. US20030170257A1
;; GENERAL INFORMATION:
;; APPLICANT: Trimell, Adama Roseanne
;; APPLICANT: Paeen, Guido Christiaan
;; APPLICANT: Nuttall, Patricia Anne
;; TITLE OF INVENTION: Vaccine Comprising A Tick Cement Protein
;; FILE REFERENCE: 2488-1-007
;; CURRENT APPLICATION NUMBER: US/10/280,114
;; PRIOR FILING DATE: 2002-10-24
;; PRIOR APPLICATION NUMBER: PCT/GB01/01834
;; PRIOR FILING DATE: 2001-04-25
;; PRIOR APPLICATION NUMBER: GB 0010068.5
;; PRIOR FILING DATE: 2000-04-25
;; PRIOR APPLICATION NUMBER: GB 0028606.2
;; PRIOR FILING DATE: 2000-11-23
;; NUMBER OF SEQ ID NOS: 17
;; SOFTWARE: Seqwin99, version 1.02
;; SEQ ID NO 9
;; LENGTH: 204
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Clone 24
US-10-280-114-9

Query Match 11.3%; Score 125; DB 12; Length 204;
Best Local Similarity 29.1%; Pred. No. 0.0043;
Matches 50; Conservative 10; Mismatches 64; Indels 48; Gaps 7;

QY 7 LPCAFLCLMHARGQDFDLADLDDEPTKKPNSDIYKPK-----KP 50
DB 11 LAFAFLLSAIIHNNQSCVDAA-----PTRRP-----MSPPGCAGPCFTGIATLLRP 59
QY 51 PYYQPEPNDGSGNITPRPKRPPQPGNSGSGYFNDVDRDGRYP-----PRP 101
DB 60 GGGQOPGGQOPGQGRP-PMRPPGPVPGTSGSPQGRPGAPR-PGPVGTSGSPQGRPNA 117
QY 102 RPRPPAGGGGGSSYGNSDNTHGGRGYPNSRYGNTYGGDHHSTYGNPEGN 153
DB 118 RPRPGPVGTPVSSPGSS-----PGSSPGISIGTPLGTPLGTPQGS 159

RESULT 15
US-10-080-170-4

;; Sequence 4, Application US/10080170
;; Publication No. US20030129601A1
;; GENERAL INFORMATION:
;; APPLICANT: COLE, S.T.
;; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
;; IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
;; TREATMENT OF MYCOBACTERIOSIS
;; FILE REFERENCE: 03495.0218
;; CURRENT APPLICATION NUMBER: US/10/080,170
;; PRIOR FILING DATE: 2002-06-10
;; PRIOR APPLICATION NUMBER: 60/270,123
;; PRIOR FILING DATE: 2001-02-22
;; NUMBER OF SEQ ID NOS: 652
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 4
;; LENGTH: 488
;; TYPE: PRT
;; ORGANISM: Mycobacterium leprae
US-10-080-170-4

Query Match 10.9%; Score 121.5; DB 16; Length 488;
Best Local Similarity 26.4%; Pred. No. 0.022;
Matches 51; Conservative 9; Mismatches 52; Indels 81; Gaps 12;

QY 33 DEPTKKPNSDIYKPKPPYTPP-----ENPDSG-----NITYR 68
DB 187 DPGQDPPRG-CYP-PKPGSYPOAGHPPLHRPDGGYDPQGGYEDORAYHDGQGGYPS 244
QY 69 PKRPPQPGNSGNSG-----GYFNDVDR-----DDGRYPPRRP 103
DB 245 PYQGRATPGYSGSHDQYRGSYGPBGQPGYGYDYGKGPAPRPGSYTTBGF 304
QY 104 RPP-----AGGGGGYS-----SYGNSDNTHR-----GYPNS-RYGN 137
DB 305 APPEQRYAYPDGGYDQGYGSHGLGYGRD--YGRQETQYVANTLPGYVAPSSGAYE 362
QY 138 TYGGDHHSTYGNP 150
DB 363 PAGRDY--DYGP 373

Search completed: February 18, 2004, 19:12:05
Job time : 36 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 18, 2004, 19:04:25 ; Search time 21 Seconds

(without alignments)
392,887 Million cell updates/sec

Title: US-09-997-701-1

Perfect score: 1110
1 MESWMLPCLAFICFLMHAR.....SYFKANRRNCFRTHPEENV 195

Sequence: BLOSUM62

Scoring table: Gap10.0 , Gape10.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCUTUS.COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1110	100.0	195	3	US-09-187-331-1
2	1110	100.0	195	4	US-09-470-946-1
3	1004.5	90.5	180	4	US-09-187-331-5
4	1004.5	90.5	180	4	US-09-470-946-5
5	118.5	10.7	264	4	US-09-431-887-24
6	117	10.5	922	3	US-08-460-269C-6
7	115.5	10.4	129	4	US-09-328-352-6408
8	115	10.4	264	3	US-09-128-450-21
9	115	10.4	264	4	US-09-823-494-21
10	114.5	10.3	643	4	US-09-196-270-3
11	112	10.1	304	4	US-09-403-343B-22
12	111.5	10.0	721	4	US-09-252-991A-28293
13	110.5	10.0	256	4	US-09-431-887-31
14	110.5	10.0	264	4	US-09-431-887-27
15	109.5	9.9	1958	1	US-07-945-283-2
16	108.5	9.8	256	4	US-09-431-887-25
17	108.5	9.8	256	4	US-09-431-887-26
18	108.5	9.8	256	4	US-09-431-887-28
19	108.5	9.8	257	4	US-09-431-887-29
20	108.5	9.8	257	4	US-09-431-887-30
21	108.5	9.8	263	1	US-08-242-188-3
22	108.5	9.8	263	1	US-08-509-251A-3
23	108.5	9.8	263	1	US-08-660-626-9
24	108.5	9.8	263	1	US-08-692-892-3
25	108.5	9.8	263	2	US-08-713-939A-3
26	108.5	9.8	263	2	US-08-868-162A-23
27	108.5	9.8	263	3	US-09-031-168-9

28	108.5	9.8	263	3	US-09-036-579-3	Sequence 3, Appl1
29	108.5	9.8	263	4	US-09-550-374-3	Sequence 3, Appl1
30	108.5	9.8	263	4	US-09-943-906-3	Sequence 3, Appl1
31	108.5	9.8	264	4	US-09-627-218B-11	Sequence 11, Appl1
32	108	9.7	256	3	US-09-128-450-22	Sequence 22, Appl1
33	108	9.7	256	4	US-09-823-494-32	Sequence 22, Appl1
34	108	9.7	891	4	US-08-714-741-32	Sequence 32, Appl1
35	107.5	9.7	911	3	US-08-460-269C-4	Sequence 4, Appl1
36	106.5	9.6	907	3	US-08-989-299-12	Sequence 12, Appl1
37	106	9.5	369	4	US-09-252-991A-20460	Sequence 20460, A
38	105.5	9.5	538	4	US-09-309-572-12	Sequence 12, Appl1
39	105.5	9.5	1312	4	US-09-554-572-26	Sequence 26, Appl1
40	105.5	9.5	1737	4	US-09-309-572-13	Sequence 13, Appl1
41	105	9.5	303	3	US-08-818-112-92	Sequence 92, Appl1
42	105	9.5	303	4	US-08-818-111-93	Sequence 92, Appl1
43	105	9.5	303	4	US-09-056-556-92	Sequence 92, Appl1
44	105	9.5	303	4	US-09-072-596-93	Sequence 93, Appl1
45	105	9.5	466	3	US-08-526-136-13	Sequence 13, Appl1

ALIGNMENTS

```
RESULT 1
US-09-187-331-1
; Sequence 1, Application US/09187331
; Patent No. 6043056
; GENERAL INFORMATION:
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Gorgone, Gina A.
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: CELL SURFACE GLYCOPROTEINS
; FILE REFERENCE: PF-0631 US
; CURRENT APPLICATION NUMBER: US/09/187,331
; CURRENT FILING DATE: 1998-11-06
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 2297891
US-09-187-331-1

Query Match      100.0%; Score 1110; DB 3; Length 195;
Best Local Similarity 100.0%; Pred. No. 3, 8e-91;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MESWMLPCLAFICFLMHARGQDFDLADLDPEPTKKNSDIYKPKPPYPOEPENPD 60
DB      1 MESWMLPCLAFICFLMHARGQDFDLADLDPEPTKKNSDIYKPKPPYPOEPENPD 60
QY      61 SGNITPRKPRPQPPQPGNSGSGYFNDVDDGPPRRPRRPAAGGGGGYSSYGS 120
DB      61 SGNITPRKPRPQPPQPGNSGSGYFNDVDDGPPRRPRRPAAGGGGGYSSYGS 120
QY      121 DNTHGGRGYPNRYNTYGGDHSYTGNEGMVAKIYSPIVSVVVTLLGAAAYFKL 180
DB      121 DNTHGGRGYPNRYNTYGGDHSYTGNEGMVAKIYSPIVSVVVTLLGAAAYFKL 180
QY      181 NNRRNCFRTHPEENV 195
DB      181 NNRRNCFRTHPEENV 195

RESULT 2
US-09-470-946-1
; Sequence 1, Application US/09470946
; Patent No. 6358923
; GENERAL INFORMATION:
```

APPLICANT: Yue, Henry
APPLICANT: Corley, Neil C.
APPLICANT: Guegler, Karl J.
APPLICANT: Gorgone, Gina A.
APPLICANT: Baughn, Mariah R.
TITLE OF INVENTION: CELL SURFACE GLYCOPROTEINS
FILE REFERENCE: PF-0631 US
CURRENT APPLICATION NUMBER: US/09/470, 946
CURRENT FILING DATE: 1999-12-22
EARLIER APPLICATION NUMBER: US 09/187, 331
EARLIER FILING DATE: 1998-11-06
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PERL Program
SEQ ID NO 1
LENGTH: 195
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE: -
OTHER INFORMATION: 2297891
US-09-470-946-1

Query Match 100.0%; Score 1110; DB 4; Length 195;
Best Local Similarity 100.0%; Pred. No. 3.8e-91;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MESWGLPCLAFCLFMAHARGQDFDLADLDDEPTKXNSDIYKPKPPYPOPENPD 60
DB 1 MESWGLPCLAFCLFMAHARGQDFDLADLDDEPTKXNSDIYKPKPPYPOPENPD 60
QY 61 SGNITPRPKRPPQPOPGNSGSGYFNDVDRDGRYPKPPRPPAGGGGGGSSYGS 120
DB 61 SGNITPRPKRPPQPOPGNSGSGYFNDVDRDGRYPKPPRPPAGGGGGGSSYGS 120
QY 121 DNTHGCGYRPNRYGNTYGGDHSTYGNPEGNVAKIVSPISVVVVTLLGAASVFKL 180
DB 121 DNTHGCGYRPNRYGNTYGGDHSTYGNPEGNVAKIVSPISVVVVTLLGAASVFKL 180
QY 181 NNRNCFRTHEPENV 195
DB 181 NNRNCFRTHEPENV 195

RESULT 3
US-09-187-331-5
Sequence 5, Application US/09187331
Patent No. 6043056
GENERAL INFORMATION:
APPLICANT: Yue, Henry
APPLICANT: Corley, Neil C.
APPLICANT: Guegler, Karl J.
APPLICANT: Gorgone, Gina A.
APPLICANT: Baughn, Mariah R.
TITLE OF INVENTION: CELL SURFACE GLYCOPROTEINS
FILE REFERENCE: PF-0631 US
CURRENT APPLICATION NUMBER: US/09/187, 331
CURRENT FILING DATE: 1998-11-06
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PERL Program
SEQ ID NO 5
LENGTH: 180
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE: -
OTHER INFORMATION: g2499136
US-09-187-331-5

Query Match 90.5%; Score 1004.5; DB 3; Length 180;
Best Local Similarity 92.3%; Pred. No. 7.8e-82;
Matches 180; Conservative 0; Mismatches 0; Indels 15; Gaps 1;

QY 1 MESWGLPCLAFCLFMAHARGQDFDLADLDDEPTKXNSDIYKPKPPYPOPENPD 60
DB 1 MESWGLPCLAFCLFMAHARGQDFDLADLDDEPTKXNSDIYKPKPPYPOPENPD 60

QY 61 SGNITPRPKRPPQPOPGNSGSGYFNDVDRDGRYPKPPRPPAGGGGGGSSYGS 120
DB 61 SGNITPRPKRPPQPOPGNSGSGYFNDVDRDGRYPKPPRPPAGGGGGGSSYGS 120
QY 121 DNTHGCGYRPNRYGNTYGGDHSTYGNPEGNVAKIVSPISVVVVTLLGAASVFKL 180
DB 121 DNTHGCGYRPNRYGNTYGGDHSTYGNPEGNVAKIVSPISVVVVTLLGAASVFKL 180
QY 181 NNRNCFRTHEPENV 195
DB 181 NNRNCFRTHEPENV 195

RESULT 4
US-09-470-946-5
Sequence 5, Application US/09470946
Patent No. 6358923
GENERAL INFORMATION:
APPLICANT: Yue, Henry
APPLICANT: Corley, Neil C.
APPLICANT: Guegler, Karl J.
APPLICANT: Gorgone, Gina A.
APPLICANT: Baughn, Mariah R.
TITLE OF INVENTION: CELL SURFACE GLYCOPROTEINS
FILE REFERENCE: PF-0631 US
CURRENT APPLICATION NUMBER: US/09/470, 946
CURRENT FILING DATE: 1999-12-22
EARLIER APPLICATION NUMBER: US 09/187, 331
EARLIER FILING DATE: 1998-11-06
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PERL Program
SEQ ID NO 5
LENGTH: 180
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE: -
OTHER INFORMATION: g2499136
US-09-470-946-5

Query Match 90.5%; Score 1004.5; DB 4; Length 180;
Best Local Similarity 92.3%; Pred. No. 7.8e-82;
Matches 180; Conservative 0; Mismatches 0; Indels 15; Gaps 1;

QY 1 MESWGLPCLAFCLFMAHARGQDFDLADLDDEPTKXNSDIYKPKPPYPOPENPD 60
DB 1 MESWGLPCLAFCLFMAHARGQDFDLADLDDEPTKXNSDIYKPKPPYPOPENPD 60
QY 61 SGNITPRPKRPPQPOPGNSGSGYFNDVDRDGRYPKPPRPPAGGGGGGSSYGS 120
DB 61 SGNITPRPKRPPQPOPGNSGSGYFNDVDRDGRYPKPPRPPAGGGGGGSSYGS 120
QY 121 DNTHGCGYRPNRYGNTYGGDHSTYGNPEGNVAKIVSPISVVVVTLLGAASVFKL 180
DB 121 DNTHGCGYRPNRYGNTYGGDHSTYGNPEGNVAKIVSPISVVVVTLLGAASVFKL 180
QY 181 NNRNCFRTHEPENV 195
DB 181 NNRNCFRTHEPENV 195

RESULT 5
US-09-431-887-24
Sequence 24, Application US/09431887
Patent No. 6534036
GENERAL INFORMATION:
APPLICANT: D-Gen Limited
TITLE OF INVENTION: BIOLOGICAL MATERIALS AND METHODS USEFUL IN THE
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PRION DISEASE
FILE REFERENCE: ICOT/P21952
CURRENT APPLICATION NUMBER: US/09/431, 887
CURRENT FILING DATE: 1999-11-02
PRIOR APPLICATION NUMBER: GB 9824091.4

PRIOR FILING DATE: 1999-11-04
 NUMBER OF SEQ ID NOS: 37
 SOFTWARE: Patentin Ver. 2.0
 SEQ ID NO 24
 LENGTH: 264
 TYPE: PRT
 ORGANISM: Bos taurus
 US-09-431-887-24

Query Match 10.7%; Score 118.5; DB 4; Length 264;
 Best Local Similarity 32.9%; Pred. No. 0.0038;
 Matches 28; Conservative 9; Mismatches 33; Indels 15; Gaps 3;

QY 72 RPOPOGNSGSGYFNDVDRDGRYPFRPP-----RPPAGGGGGYSSYSGNSDNTHGRC 127
 DB 25 KRPKPKGGGNTG-----SRYPGSGSPGGRYPGSGGSGGQPHGGGQPHGGG 75

QY 128 GYRPNRYGNTYGGDHHSTYGNPEG 152
 DB 76 GW--GGPHGGGQPHGGGQPHG 98

RESULT 6
 US-08-460-269C-6
 Sequence 6; Application US/08460269C
 Patent No. 6197548
 GENERAL INFORMATION:
 APPLICANT: CLARE, JEFFREY J.
 ROMANOS, MICHAEL A.

TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN YEAST
 NUMBER OF SEQUENCES: 17
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
 STREET: 2200 Clarendon Blvd., Suite 1400
 CITY: ARLINGTON
 STATE: VA
 COUNTRY: USA
 ZIP: 22201

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/460,269C
 FILING DATE: 02-Jun-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Lebovitz, Richard M.
 REGISTRATION NUMBER: 37,067
 REFERENCE/DOCKET NUMBER: Popov-2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 243-6333
 TELEFAX: (703) 243-6410

INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 922 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-08-460-269C-6

Query Match 10.5%; Score 117; DB 3; Length 922;
 Best Local Similarity 27.5%; Pred. No. 0.022;
 Matches 39; Conservative 9; Mismatches 50; Indels 44; Gaps 5;

QY 15 FILHAAGQRFDLADLDDEPTKKNSDIYKPKPPYQFNPDSGNTYFRPPRPO 74
 DB 550 YRLAANGGWSLVGAAPAPAPAPQPGQPGQPPQPPQPP-----PQPPQPP 600

QY 75 PPGNSGSGYFNDVDRDGRYPFRPPRPPAG-----GGGGYSS---YGN 119

DB 601 PQ-----RQEPAPQPPAGBELSAANAAVNTGVGLASTLWAE 641
 QY 120 SDNTHGR--GYRPNRYGNTY 140
 DB 642 SNLSRGLGELRLNPDAGAWG 663

RESULT 7
 US-09-328-352-6408
 Sequence 6408; Application US/09328352
 Patent No. 6562958
 GENERAL INFORMATION:
 APPLICANT: Gary L. Breton et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 FILE REFERENCE: GTC99-03PA
 CURRENT APPLICATION NUMBER: US/09/328,352
 CURRENT FILING DATE: 1999-06-04
 NUMBER OF SEQ ID NOS: 8252
 SEQ ID NO 6408
 LENGTH: 129
 TYPE: PRT
 ORGANISM: Acinetobacter baumannii
 US-09-328-352-6408

Query Match 10.4%; Score 115.5; DB 4; Length 129;
 Best Local Similarity 30.1%; Pred. No. 0.0031;
 Matches 25; Conservative 12; Mismatches 21; Indels 25; Gaps 3;

QY 34 PEPTKKPNSD-----IYKPKPPYQFNPDSGNTYFRPP--RPOPOGNSGS 83
 DB 50 PRPFRPRSSSSSSSSRRLRPRPRPRSSSSSSSSRRLRPRPRPRSSSSSSSR 109

QY 84 GGYFNDVDRDGRYPFRPPRPP 106
 DB 110 -----LRPRPRPP 117

RESULT 8
 US-09-128-450-21
 Sequence 21; Application US/09128450
 Patent No. 6211149
 GENERAL INFORMATION:
 APPLICANT: Chesebro, Bruce W
 APPLICANT: Caughey, Byron W
 APPLICANT: Chadrey, Joelle
 APPLICANT: Priola, Susette
 TITLE OF INVENTION: Inhibitors of Formation of Protease Resistant Prion
 FILE REFERENCE: 50121
 CURRENT APPLICATION NUMBER: US/09/128,450
 CURRENT FILING DATE: 1998-08-03
 NUMBER OF SEQ ID NOS: 29
 SOFTWARE: Patentin Ver. 2.0
 SEQ ID NO 21
 LENGTH: 264
 TYPE: PRT
 ORGANISM: Bos taurus
 US-09-128-450-21

Query Match 10.4%; Score 115; DB 3; Length 264;
 Best Local Similarity 32.9%; Pred. No. 0.0078;
 Matches 28; Conservative 9; Mismatches 32; Indels 16; Gaps 3;

QY 72 RPOPOGNSGSGYFNDVDRDGRYPFRPP-----RPPAGGGGGYSSYSGNSDNTHGRC 127
 DB 25 KRPKPKGGGNTG-----SRYPGSGSPGGRYPGSGGSGGQPHGGGQPHGGG 75

QY 128 GYRPNRYGNTYGGDHHSTYGNPEG 152
 DB 76 WQGP---HGCGWQPHGGGQPHG 97

RESULT 9
US-09-823-494-21
Sequence 21, Application US/09823494
Patent No. 655610
GENERAL INFORMATION:
APPLICANT: Chesebro, Bruce W
APPLICANT: Caughey, Byron W
APPLICANT: Chabry, Joelle
APPLICANT: Priola, Susanne
TITLE OF INVENTION: Inhibitors of Formation of Protease Resistant Prion
FILE REFERENCE: 50121
CURRENT APPLICATION NUMBER: US/09/823,494
CURRENT FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: 09/128,450
PRIOR FILING DATE: 1998-08-03
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 21
LENGTH: 264
TYPE: PRT
ORGANISM: Bos taurus
US-09-823-494-21

Query Match 10.4%; Score 115; DB 4; Length 264;
Best Local Similarity 32.9%; Pred. No. 0.0078;
Matches 28; Conservative 9; Mismatches 32; Indels 16; Gaps 3;
QY 72 RPOGPNNSGSGYFNDVDRDGRYPRRP-----RPPGCGGCGGYSYSGNSDNTHTG 127
DB 25 KKRKPKPGGKMTGG-----SRYPGQSGPGGNRYPPQGSGGQPHGSGMGQPHGG 75
QY 128 GYRPNRSRYGNTYGGDHSHTYGNPEG 152
DB 76 WGP-----HGGGWGPFGGGMQPHG 97

RESULT 10
US-09-196-270-3
Sequence 3, Application US/09196270
Patent No. 650636
GENERAL INFORMATION:
APPLICANT: Hecht, Peter
APPLICANT: Madden, Kevin
TITLE OF INVENTION: CHIMERIC PRE-ACTIVATED TRANSCRIPTION
FILE REFERENCE: 50078/004002
CURRENT APPLICATION NUMBER: US/09/196,270
CURRENT FILING DATE: 1998-11-19
EARLIER APPLICATION NUMBER: 60/066,129
EARLIER FILING DATE: 1997-11-19
EARLIER APPLICATION NUMBER: 60/066,308
EARLIER FILING DATE: 1997-11-21
EARLIER APPLICATION NUMBER: 60/066,462
EARLIER FILING DATE: 1997-11-24
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 643
TYPE: PRT
ORGANISM: Penicillium chrysogenum
US-09-196-270-3

Query Match 10.3%; Score 114.5; DB 4; Length 643;
Best Local Similarity 20.7%; Pred. No. 0.024;
Matches 55; Conservative 20; Mismatches 58; Indels 133; Gaps 12;
QY 34 PEPTKNSDIYPR-----EKPPY-----PQENPDG 62
DB 155 PEPGKHDMFMFPQNPGRSPATHYFESPINGINGQYSHAPPPQYQPHPPQAPNPHSY 214
QY 63 GNTPRPKPRRPPQPGNSGNSG-----YFNVDR-----DDGRY----- 98

DB 215 GNLY-----YALSGQEGGHPYDRKRGYDALNEFFGDKRQFDNSYAAGQRL 265
QY 99 -----PRPPA-GGGGGYSS----- 116
DB 266 GIALQLPPLSGPAPRYQMPAPVAVGGGGGCGGAPAPPGYHLPRMNNVTKNDLNI 325
QY 117 -----YNSDNTHTGGRGYRPNRSRYGNTYGG-DHSHTYGNPEGMVAIV----- 159
DB 326 DQFLBQMNTIYESDENVAAGVQAQAHY--VHGGMNRTTHSPPTHSRQATLLQPSA 383
QY 160 -----SPISVVVVTLLGAASY 177
DB 384 PMAAATASHPSVGTPTALTPSSAQS 409

RESULT 11
US-09-403-343B-22
Sequence 22, Application US/09403343B
Patent No. 6555091
GENERAL INFORMATION:
APPLICANT: JOLIVET-RENAUD, COLETTE
APPLICANT: PERRON, HERVE
APPLICANT: MANDRAND, BERNARD
TITLE OF INVENTION: POLYPEPTIDE CAPABLE OF REACTING WITH ANTIBODIES OF
FILE REFERENCE: 104574
CURRENT APPLICATION NUMBER: US/09/403,343B
CURRENT FILING DATE: 1999-10-18
PRIOR APPLICATION NUMBER: FR/97/05679
PRIOR FILING DATE: 1997-04-29
PRIOR APPLICATION NUMBER: FR/97/16870
PRIOR FILING DATE: 1997-12-31
PRIOR APPLICATION NUMBER: PCT/FR98/00870
PRIOR FILING DATE: 1998-04-29
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 22
LENGTH: 304
TYPE: PRT
ORGANISM: MSRV
US-09-403-343B-22

Query Match 10.1%; Score 112; DB 4; Length 304;
Best Local Similarity 31.0%; Pred. No. 0.017;
Matches 35; Conservative 13; Mismatches 39; Indels 26; Gaps 7;
QY 13 LCFM-----HARGQRFDLADLDPEPTKPNDSIYPRKPYQPE-NPSSGNTY 67
DB 73 LCLLSLAPHQHGHQ-----DKQTQTPPPPP-----PPQPLTPRPDANPSI--NSHN 121
QY 68 RPKRPQPPGN-----SGNSGGYFNDVDRDGRY-----PRRPRPPAGG 110
DB 122 KPKNEBGTDGHAQGRKRTKGPDPDPGSGFYLKTLPPPPPPPTGPG 174

RESULT 12
US-09-252-991A-28293
Sequence 28293, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 28293


```

; LENGTH: 721
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28293

Query Match
Best Local Similarity 10.0%; Score 111.5; DB 4; Length 721;
Matches 50; Conservative 16; Mismatches 63; Indels 49; Gaps 11;

QY 39 KPNSDIYKPKPPYPO-----PENPDSGNIVPRPKP-----RPOP-----75
DB 121 RPAADRQLPQPPRRPPYPPGPFPTGTGTGRLARSPGGRACGTGAPGQPRPDGLD 180
QY 76 -QPNSSGSGCYENDVDNRD--DGRYP-PRPRPPAGGGGGYSSYNSDNTGRCGYRP 131
DB 181 RQPHRVPSSGHLQDGPBPAPGRRPVAPARAPAPAG-----RCVHADADH-RQYQRP 234
QY 132 NSRGNVTYGGDHSHTVG-NPEGNVAKIVSPYVVVVTLLGAASFKANRNCRR 188
DB 235 -----GDHRRRGRRPDPROAG--APCTARATT--GRAGSGRLTSRPGCRR 277

RESULT 13
US-09-431-887-31
; Sequence 31, Application US/09431887
; Patent No. 6534036
; GENERAL INFORMATION:
; APPLICANT: D-Gen Limited
; TITLE OF INVENTION: BIOLOGICAL MATERIALS AND METHODS USEFUL IN THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PRION DISEASE
; FILE REFERENCE: ICOT/P21952
; CURRENT APPLICATION NUMBER: US/09/431,887
; CURRENT FILING DATE: 1999-11-02
; PRIOR APPLICATION NUMBER: GB 9824091.4
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 31
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Canis familiaris
US-09-431-887-31

Query Match
Best Local Similarity 10.0%; Score 110.5; DB 4; Length 256;
Matches 44; Conservative 12; Mismatches 67; Indels 21; Gaps 6;

QY 70 KPRPOPQSGNSGSGYFNVDNDDGRYPRPRP--RPPAGGGGGYSSYNSDNTGRCG 127
DB 24 KRRPKRGKMTGGGSRTPGGSGPGRNRYPPQGGGQPHGGGQPHGGGQPHGGG 83
QY 128 GYRPNRYGNTYGGDHSHTVG--NPEGNV---AKIVSPYVVVVTLLGAAS----176
DB 84 WGGPHGG-GWGGGSGSHSGWKPKTKMKVHAGAAAAGAVVGGVGLGSGMSMRPL 132
QY 177 -----YFKANRNCRRCTHREPNV 195
DB 143 HFGNDYEDRYRREMYR--YPEQV 164

RESULT 14
US-09-431-887-27
; Sequence 27, Application US/09431887
; Patent No. 6534036
; GENERAL INFORMATION:
; APPLICANT: D-Gen Limited
; TITLE OF INVENTION: BIOLOGICAL MATERIALS AND METHODS USEFUL IN THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF PRION DISEASE
; FILE REFERENCE: ICOT/P21952
; CURRENT APPLICATION NUMBER: US/09/431,887
; CURRENT FILING DATE: 1999-11-02
; PRIOR APPLICATION NUMBER: GB 9824091.4
; PRIOR FILING DATE: 1999-11-04
```

```

; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 27
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Tragelaphus strepsiceros
US-09-431-887-27

Query Match
Best Local Similarity 10.0%; Score 110.5; DB 4; Length 264;
Matches 27; Conservative 9; Mismatches 34; Indels 15; Gaps 3;

QY 72 RPOPQSGNSGSGYFNVDNDDGRYPRPRP--RPPAGGGGGYSSYNSDNTGRCG 127
DB 25 KRRPKRGKMTGG-----SRYPGQSGPGRNRYPPGQGGGQPHGGGQPHGGG 75
QY 128 GYRPNRYGNTYGGDHSHTYNGEG 152
DB 76 GW--GQPHGGGQPHGGGQPHG 98

RESULT 15
US-07-945-283-2
; Sequence 2, Application US/07945283
; Patent No. 5352596
; GENERAL INFORMATION:
; APPLICANT: Cheung, Andrew K.
; APPLICANT: Wealey, Ronald D.
; TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
; TITLE OF INVENTION: Involving The EPO and LIT Genes
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Curtis P. Ribando
; STREET: 1815 No. 5352596th University Street
; CITY: Georgia
; STATE: IL
; COUNTRY: USA
; ZIP: 61604
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/945,283
; FILING DATE: 19920911
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ribando, Curtis P
; REGISTRATION NUMBER: 27976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 309-685-4128
; TELEFAX: 309-685-4011 ext.513
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1958 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: 1linear
; MOLECULE TYPE: protein
US-07-945-283-2

Query Match
Best Local Similarity 9.9%; Score 109.5; DB 1; Length 1958;
Matches 42; Conservative 5; Mismatches 22; Indels 83; Gaps 9;

QY 48 PPKPYPOPENPDSGNIVP-----RPKR-POPQ-----77
DB 393 PPRP--PRSTSSSSSHQBPSTPRPPRPPRPPRPPRPPRPPRPPRPPRPPRPP 450
QY 78 -----GNSGNSGSGYFNVDNDDGRYPRPRP-----P 105
DB 451 SNNKLFSHPMGEGGDDGTAG--EGDRDPR--PSPRPPRPPRPPRPPRPPRPPRPP 506
```

Qy 106 PAGG-----GGGGSSYGNSDNTHGRCYR 130
Db 507 PAGGARRRRRGGG-----GPGRGRR 530
Search completed: February 18, 2004, 19:07:37
Job time : 22.86cs